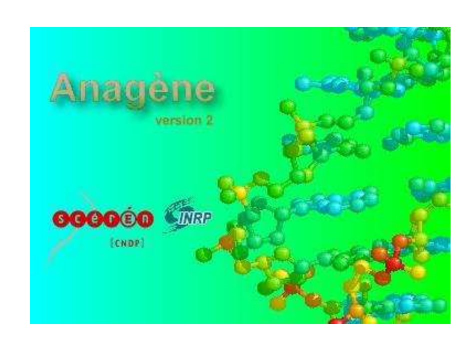


Remerciements

Cette publication emprunte aux documentations précédentes des logiciels SEQAIDII et Anagène, ainsi qu'aux multiples dossiers constitués par les équipes d'enseignants associés à l'INRP, sur les sites Internet Biogéo et Biotic consultables aux adresses suivantes :

http://www.inrp.fr/Acces/Biogeo
http://www.inrp.fr/Acces/biotic

Y ont contribué, notamment, par les contacts avec les spécialistes, la synthèse des informations scientifiques, la sélection des séquences moléculaires, la conception, l'expérimentation et la suggestion de démarches pédagogiques :

Jacques Barrère, Lycée Paul-Louis Courrier, TOURS
Jean-Yves Dupont, IA-IPR ORLEANS-TOURS
Monique Dupuis, IA-IPR NANTES
Jean-Claude Hervé, IA-IPR VERSAILLES
Anne Florimond, Lycée Richelieu, RUEIL-MALMAISON
Françoise Jauzein, INRP
Alain Pothet, Lycée Marie Curie, SCEAUX
Nadia Ouahioun, Lycée Montesquieu, LE-PLESSIS-ROBINSON
Bernard Therrié, Lycée Parc des Loges, EVRY
Vincent Thizeau, Lycée Louis Bascan, RAMBOUILLET

Collection « Analyse de données, simulation… »

ANAGÈNE

Analyse de séquences nucléiques et protéiques

Version 2006 du logiciel et de la banque de données

Sélection en sciences de la vie et de la Terre

À consulter dans les logithèques des CRDP et CDDP
et à découvrir dans la librairie sur le site du SCÉRÉN-CNDP
http://www.cndp.fr

COLLECTION « TEXTES DE RÉFÉRENCE »

SVT	Programmes, accompagnement et autres ressources

COLLECTION « ANALYSE DE DONNÉES, SIMULATION… »

Anagène	Analyse de séquences nucléiques et protéiques
En Jeu Ma Planète	Découverte ludique des enjeux de l'environnement
Génie génétique II	Simulation d'une manipulation génétique
Géocéan : vol. 1, 2, 3	Géodynamique océanique, subduction et expansion
Léman	Un écosystème en chiffres
Magma	Simulation de la cristallisation magmatique
Nut	Équilibre de la ration alimentaire
Réhor	Régulation de la sécrétion des hormones sexuelles
Rouge-gorge	Étude expérimentale d'un comportement

COLLECTION « IMAGERIE SCIENTIFIQUE »

Explorer le corps	Surfaces d'échange
Explorer le corps	Communication nerveuse
Explorer la Terre	La télédétection

COLLECTION « DÉCOUVERTE DE LA NATURE »

Les Oiseaux d'Europe	Encyclopédie multimédia interactive permettant de gérer ses observations
Découvrir la planète Terre	Notre environnement en 7 thèmes : la planète, l'eau, l'air, la Terre, la vie, les Hommes, les défis

COLLECTION « PROGRÈS »

La Forêt	Un peuplement de feuillus
La Friche	Un milieu riche
L'Étang	Découvrir un écosystème
Infogène	Ressources pour l'étude de l'information génétique
Communication et systèmes nerveux	Explorer, identifier, analyser des documents
Mitose-Méiose	Conservation et brassage de l'information génétique

HORS COLLECTIONS

Aléas et Enjeux	Éduquer pour prévenir les risques majeurs

Collection « Analyse de données, simulation… »

ANAGÈNE

Analyse de séquences nucléiques et protéiques

Version 2006 du logiciel et de la banque de données

Centre national de documentation pédagogique
Institut national de recherche pédagogique

L'édition de logiciels et multimédias
est coordonnée par le CNDP
Service national des productions imprimées et numériques
Téléport 1 @4 – BP 80158 86961 Futuroscope cedex
Chef de service :
Jean-François Timmel

Chefs de projet	Maurice Bouyssou (coordination) Catherine Bouyssou Daniel Buret Caroline d'Atabekian Marie-Line Périllat-Mercerot
Développement	Robert Lagoutte Yves Lévine
Secrétariat de rédaction	Pascale Langlois
Micro-édition	Fabien Biglione Sylvie Morel Ghislaine Rochette
Graphisme	Jacques Auclerc-Galland
Fabrication	Thierry Lacorne

ISBN 2-240-02523-9 (3e édition, © CNDP-INRP, 2006)
ISBN 2-240-00530-0 (2e édition, © CNDP-INRP, 2003)
ISBN 2-240-00479-7 (1re édition, © CNDP-INRP, 1997)
ISSN : 1284-1285

Collection « Analyse de données, simulation… »

ANAGÈNE

Analyse de séquences nucléiques et protéiques

Discipline Biologie

Niveau Lycée, enseignement supérieur

Sujet Universalité et variabilité de l'information génétique ; Parenté et diversité des organismes ; Des phénotypes à différents niveaux d'organisation du vivant ; Parenté entre êtres vivants actuels et fossiles – Phylogenèse – Évolution ; Stabilité et variabilité des génomes et évolution ; Procréation ; Immunologie ; Des débuts de la génétique aux enjeux actuels des biotechnologies.

Conception CNDP, Service national des productions imprimées et numériques,
Maurice Bouyssou,
professeur de sciences naturelles.
INRP, équipe de recherche technologique – éducation ACCES,
Naoum Salamé,
ingénieur de recherche.

Documentation scientifique et pédagogique Jean-Claude Hervé,
IA-IPR de sciences de la vie et de la Terre.
Monique Dupuis,
IA-IPR de sciences de la vie et de la Terre.

Adaptation et développement informatique CNDP, Service national des productions imprimées et numériques,
Érik Boucher de Crèvecoeur, ingénieur en informatique et biologie,
Monique Blavin, Jean-Jacques Équoy, Bertrand Guyonnet.

Aide en ligne Maïté Defaysse, Yvan Vincent,
bio-informaticiens.

Adaptation de l'aide et de la documentation à la version 1.8 Jean-Claude Le Hir,
professeur de sciences naturelles et formateur académique

Tests et calculs de matrices Alexandre Skiada,
élève stagiaire au CNDP.

Conduite du projet Maurice Bouyssou,
professeur de sciences naturelles et d'option informatique,
coordinateur des collections hors ligne au CNDP/SnPIN.

Configuration minimale

Compatible PC *Pentium*, 32 Mo de RAM, carte vidéo milliers de couleurs,
Windows 98 et suivants, 40 Mo requis pour l'installation complète.

SOMMAIRE

PRÉSENTATION

Anagène est un logiciel d'analyse de séquences nucléiques et protéiques qui propose quelques-uns des traitements présents dans les logiciels :

- *Seqaid II* (D.D. Rhoads et D.J. Rouffa, université du Kansas, © 1987, 1989, 1992 pour *MS-DOS*) qui comporte une gamme de traitements simples pour l'édition et l'analyse de séquences nucléiques et protéiques ;
- *Clustalw* (D. Higgins, European Molecular Biology Laboratory [EMBL], version 1.8, © 1998, pour *MS-DOS*), qui est spécialisé dans l'alignement multiple de séquences.
- *Antheprot* (G. Deléage, Institut de biologie et de chimie des protéines à Lyon, © 1995-2002), qui contient une fonction de comparaison de deux séquences entre elles produisant un graphique de ressemblance (DotPlot).

Cette version 2 d'*Anagène,* diffusée sur cédérom, a été totalement réécrite pour s'adapter à tous les environnements *Windows* et apporter des améliorations d'ordres fonctionnel et ergonomique, notamment en ce qui concerne l'accès aux thèmes d'étude, la puissance des outils de traitement, la navigation dans les séquences et le repérage des positions des éléments, la gestion des nouveaux formats de médias dont ceux accessibles en ligne.

D'autre part, cette nouvelle version est enrichie d'une banque de séquences constituée en fonction des programmes en vigueur et totalement intégrée au logiciel. Cette banque regroupe toutes les séquences nécessaires pour réaliser les propositions d'activités actuellement accessibles en ligne sur le site *Biotic* de l'INRP. On notera en particulier la présence de plusieurs thèmes d'étude destinés à la classe de seconde, de multiples exemples adaptés à l'étude des relations entre génotype, phénotype et environnement en classe de première et aux différents chapitres de terminale S qui concernent l'évolution, l'immunologie et la procréation.

Il convient également de relever que nombre de ces thèmes d'étude sont susceptibles de s'appuyer sur la visualisation des molécules en trois dimensions. On trouvera donc, en relation avec ces thèmes, les fichiers .PDB nécessaires qui peuvent être visualisés avec les logiciels *Rasmol* ou *RasTop*, tous les deux téléchargeables gratuitement.

Certains thèmes peuvent également se prêter à des prolongements en termes d'établissement des relations de parenté entre les organismes impliqués. Les données moléculaires nécessaires sont disponibles par ailleurs avec le logiciel *Phylogène* qui est également téléchargeable gratuitement.

Rappelons enfin que les prochaines mises à jour du logiciel et de sa documentation seront téléchargeables à partir du site du CNDP :

http://www.cndp.fr/svt/anagene/

qui donne également à l'adresse des mises à jour des banques de données sur le site du l'INRP :

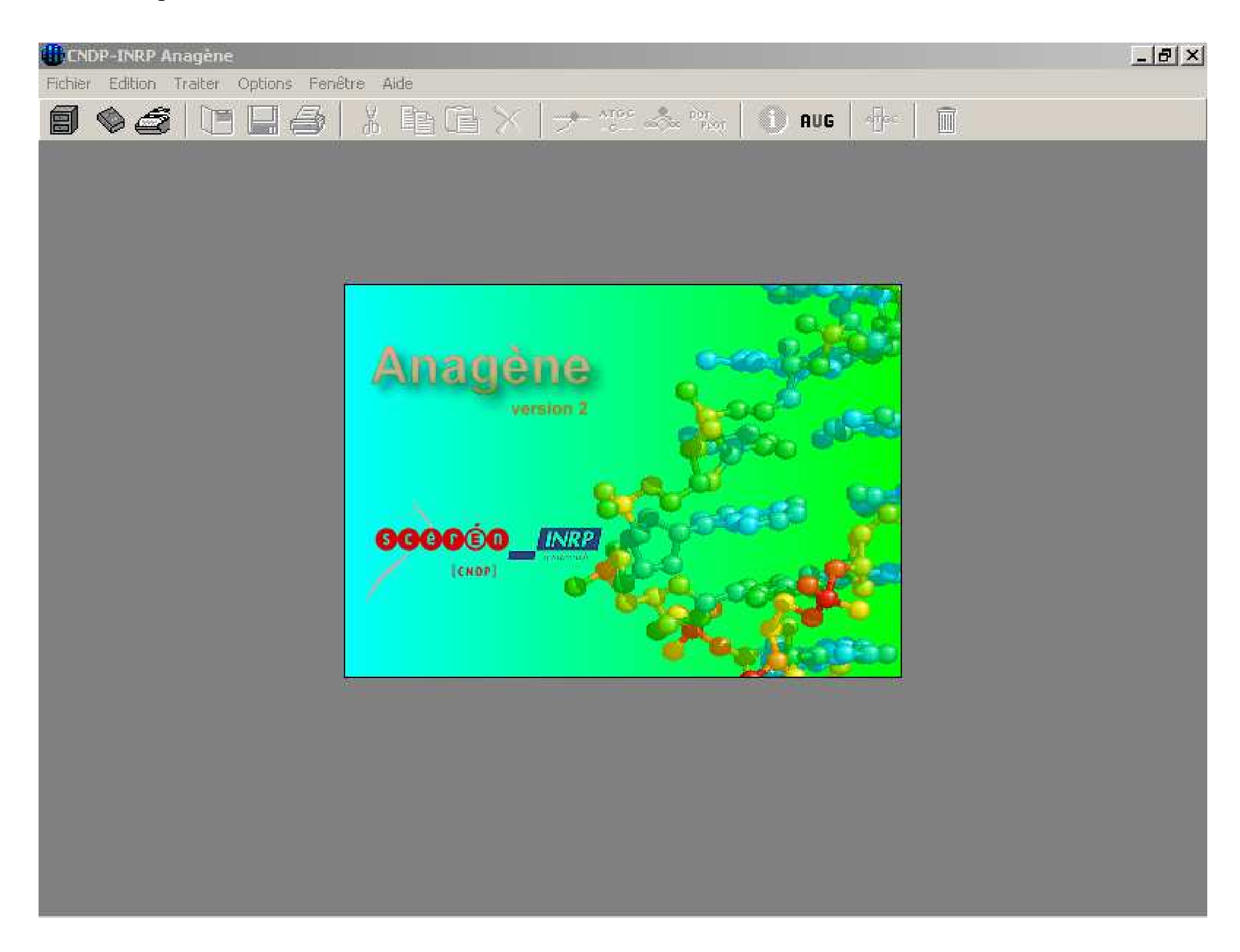
http://www.inrp.fr/Acces/biotic/

La présente documentation emprunte à deux publications de l'INRP — *Analyse de séquences nucléiques et protéiques avec le logiciel Seqaid II* (1993) et *Thèmes d'étude en génétique humaine* (1995) — ainsi qu'aux dossiers scientifiques et pédagogiques consultables en ligne sur le site *Biotic* de l'INRP.

L'ENVIRONNEMENT D'*ANAGÈNE*

Anagène fonctionne dans l'environnement *Windows*. Les dimensions de la fenêtre d'exécution du programme s'adaptent automatiquement aux paramètres d'affichage définis dans la configuration. Néanmoins, pour améliorer le confort visuel, il est préférable de travailler avec une résolution de 800 x 600 pixels au minimum. On se référera au manuel d'utilisation de *Windows* ou à celui de la carte graphique pour effectuer les adaptations nécessaires.

Sous *Windows 95* et versions suivantes, lors de l'installation, *Anagène* est ajouté à la liste des programmes accessibles par le menu **Démarrer.**

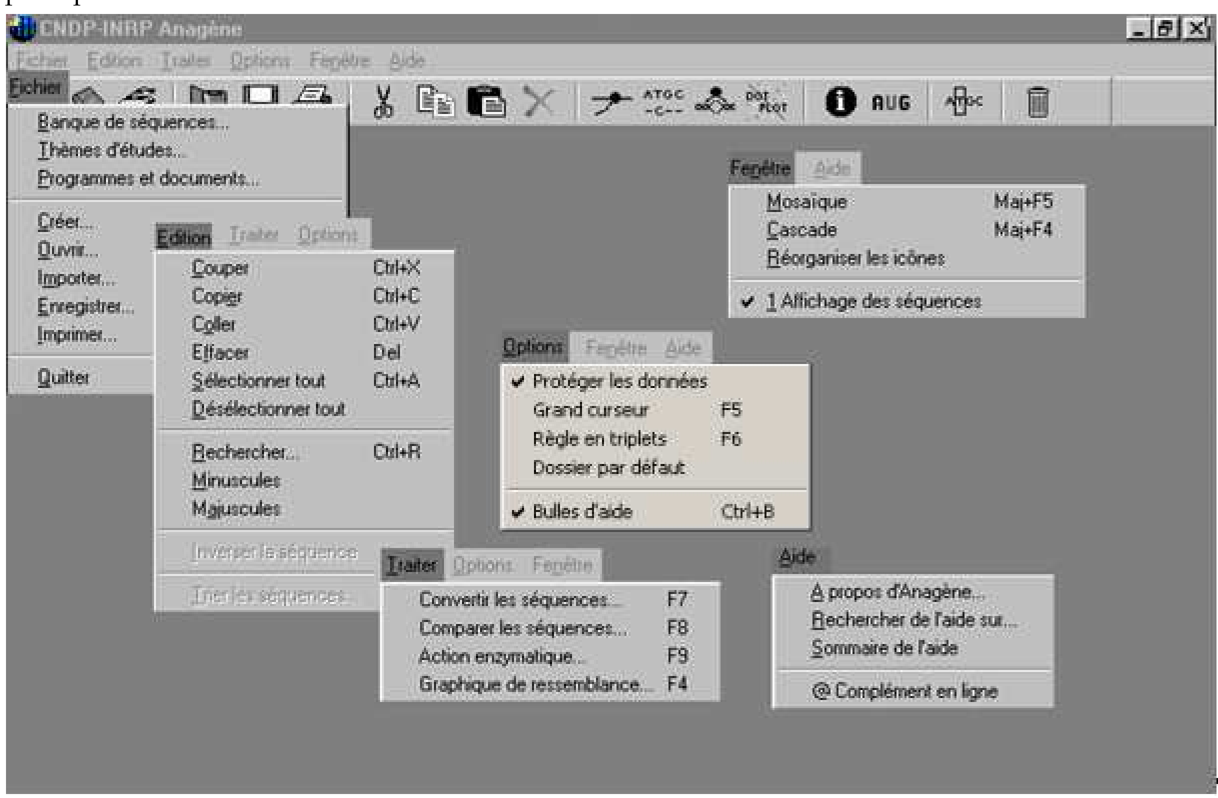


Après un temps de chargement d'autant plus rapide que le micro-ordinateur est de type récent et une phase d'initialisation nécessaire, l'interface d'*Anagène* apparaît.

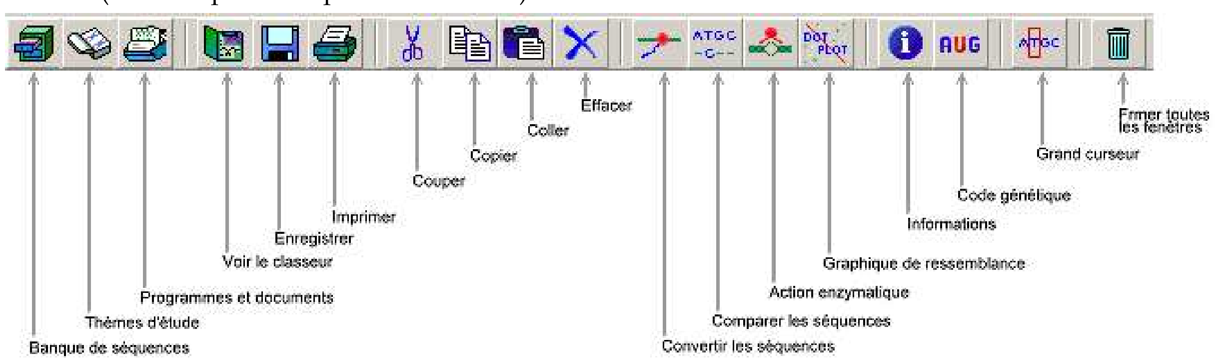
L'INTERFACE D'*ANAGÈNE*

Anagène est constitué d'une fenêtre-mère présentant, sous la barre de titre, une barre de menus et une barre d'outils.

Les modes de déplacement et de sélection pilotés à l'aide de la Souris ou du clavier obéissent aux règles adoptées par *Windows*.



Par ailleurs, lorsque le pointeur de la Souris est maintenu quelques instants en place sur l'un des objets de l'écran, tel par exemple, un bouton de la barre d'outils, une bulle d'aide apparaît et rappelle la fonction qui lui est associée (si cette option n'a pas été décochée).



Ainsi les fonctions d'*Anagène* sont accessibles soit à travers les différentes commandes de la barre des menus, soit directement pour la plupart par l'intermédiaire des boutons de la barre d'outils. Certaines commandes des menus sont aussi directement applicables par l'utilisation de raccourcis clavier ; le détail en est précisé dans le Guide de référence (page 29 et suivantes) et accompagne les commandes dans les menus déroulants du logiciel.

Menu Fichier

On y trouve les options classiques de gestion de fichiers, à savoir l'accès aux banques de données du logiciel (séquences, thèmes d'étude, programmes et documents), la création, l'ouverture, l'importation, l'enregistrement, l'impression de fichiers de séquences, et la possibilité de quitter le logiciel.

Menu Édition

Il permet d'accéder aux commandes présentes dans la plupart des logiciels d'édition ou de traitement de texte, à savoir les commandes liées au presse-papiers de *Windows* (**Couper**, **Copier**, **Coller**), les commandes d'effacement, de sélection, de recherche, d'inversion et de tri de séquences. Certaines de ces commandes ne sont applicables aux séquences que si la commande **Protéger les données** du menu **Options** n'est plus cochée. En effet, au lancement du logiciel, les séquences chargées dans la fenêtre d'édition sont par défaut ainsi protégées.

Menu Traiter

Il contient les commandes assurant les quatre grands types de traitements qu'il est possible d'effectuer sur les séquences, à savoir :

- les conversions de séquences nucléiques en séquences peptidiques et inversement, pour les séquences peptidiques, l'affichage des différents codons possibles ;
- les comparaisons de séquences, qu'elles soient simples ou par alignement mettant en évidence les discontinuités ;
- l'action d'enzymes de restriction sur des séquences d'ADN ;
- le graphique de ressemblance (ou dotplot), représentation graphique en forme de matrice de points permettant de comparer deux séquences pour détecter les régions à forte similarité.

Menu Options

Il permet de déprotéger les séquences et d'en autoriser la modification, de définir le type de curseur (simple ou grand curseur) et de règle graduée (en caractères ou en triplets) utilisés, de définir le chemin du dossier de sauvegarde (qui devra disposer de droits en lecture et écriture) et d'afficher ou non les bulles d'aide.

Menu Fenêtre

On y retrouve de classiques options propres à *Windows*, à savoir les commandes d'organisation des fenêtres et des icônes. La liste des fenêtres-filles présentes dans la fenêtre-mère d'*Anagène* y est aussi accessible.

Menu Aide

Il contient la commande **A propos d'*Anagène*** qui renseigne sur la version du logiciel et sur son édition. L'aide en ligne accessible avec la touche de fonction F1 est aussi disponible dans ce menu. Si *Anagène* fonctionne sur un micro-ordinateur connecté à Internet, des informations et données complémentaires disponibles sur les sites Internet du CNDP et de l'INRP peuvent être consultées dans un navigateur et téléchargées. Un lien mène également vers un site dédié à *Anagène*. L'accès à ces sites est réalisé à partir de la commande **@ Complément en ligne**.

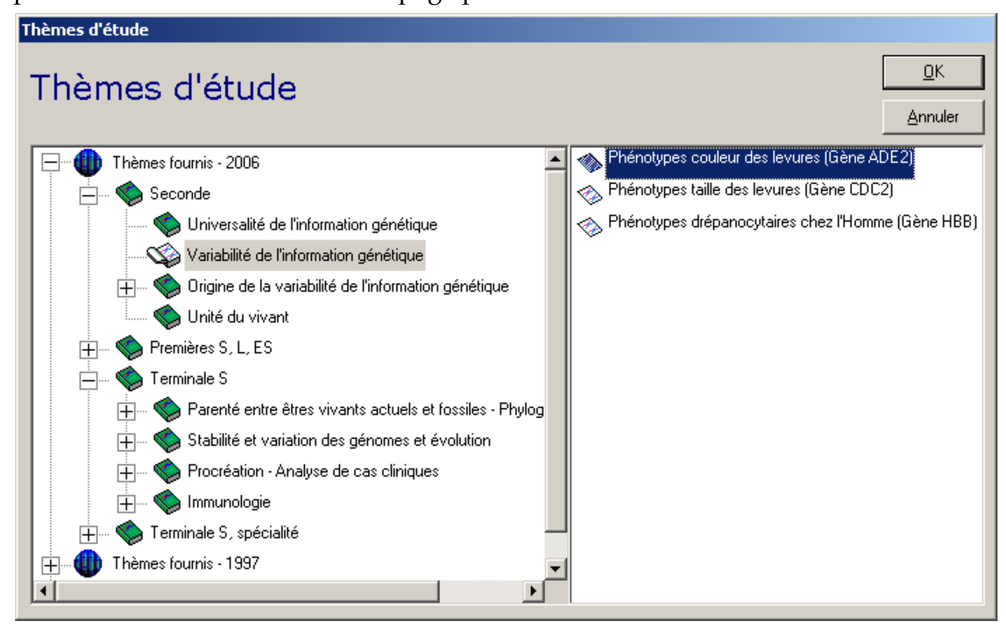
LES DONNÉES FOURNIES AVEC *ANAGÈNE*

Les thèmes d'étude

C'est l'entrée normale pour les élèves.

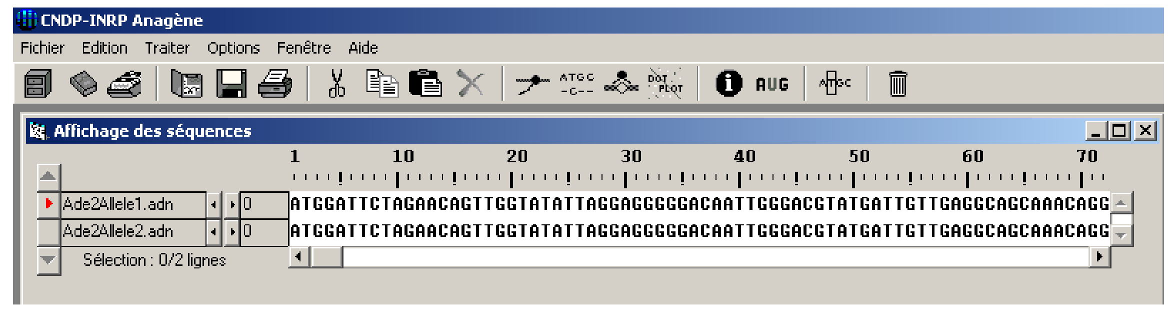
Dans la nouvelle banque de thèmes d'étude (Thèmes fournis - 2006), ceux-ci sont organisés par niveau. Les thèmes de la précédente version d'*Anagène* (Thèmes fournis - 1997) ont été maintenus après la nouvelle banque.

Les thèmes sont regroupés par grandes sections correspondant à différents contenus d'apprentissage en relation avec les programmes. Dans chaque section, on trouve les problématiques introduites par chaque groupe de séquences et si nécessaire un découpage plus fin en thèmes d'étude.



Les séquences sont regroupées et classées en dossiers thématiques, eux-mêmes ordonnés de manière arborescente. Pour les atteindre, il convient de déployer plus ou moins cette arborescence et de sélectionner le dossier désiré ; il n'est possible de sélectionner qu'un seul thème d'étude à la fois. Dans la partie droite de la boîte de dialogue s'affiche alors la liste des groupes de séquences disponibles dans ce dossier. Le bouton **Annuler** de la boîte de dialogue provoque sa fermeture et le thème qui a pu être sélectionné n'est pas chargé. La validation de la sélection d'un ou de plusieurs groupes de séquences d'un même dossier thématique s'obtient, ici encore, en cliquant sur le bouton **OK** ou en appuyant sur la touche **Entrée.** Cette validation provoque la fermeture de la boîte de dialogue et le chargement de l'ensemble des séquences que le thème d'étude peut contenir. La fenêtre d'**Affichage des séquences** apparaît alors à l'écran.

NB : il est conseillé de ne pas maximiser les dimensions de la fenêtre d'édition de séquences si l'on souhaite ouvrir également des documents du thème car cette propriété sera héritée par la fenêtre d'affichage des documents, ce qui peut entraîner des distorsions.



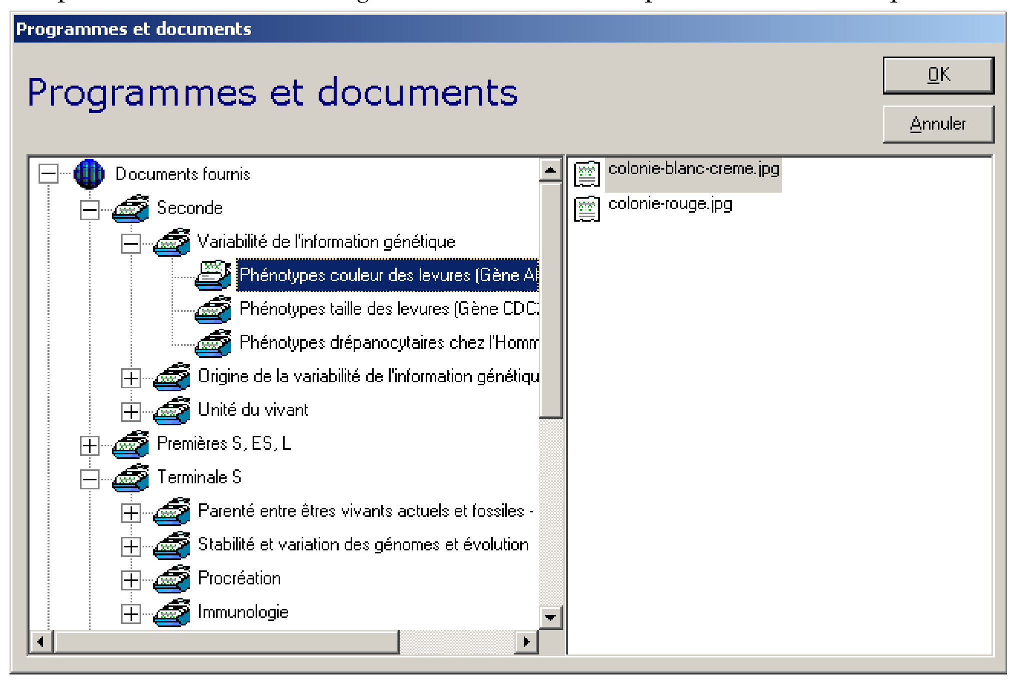
La répétition de cette procédure permet de charger d'autres thèmes d'étude. Un message signalant le changement de thème d'étude demande alors s'il n'est pas souhaitable de fermer les autres fenêtres ouvertes lors de l'étude du thème précédent de façon à vider totalement l'espace de travail avant de charger un nouveau thème d'étude.

Les programmes et documents

À chaque thème d'étude, une série de documents peut être associée (images fixes, tableaux, schémas, séquences animées). Il appartient à l'enseignant de les exploiter dans l'ordre qui lui convient.

L'exploration de l'arborescence des **Programmes et documents** disponibles est semblable à celle des autres banques. Le libellé des titres sélectionnés apparaît toujours en inversion vidéo et les boutons **OK** et **Annuler** exercent les fonctions déjà rencontrées dans les paragraphes précédents.

Dans la nouvelle banque de documents fournis, ceux-ci sont organisés par niveau et par thème d'étude. Les documents de la précédente version d'*Anagène* ont été maintenus après la nouvelle banque.



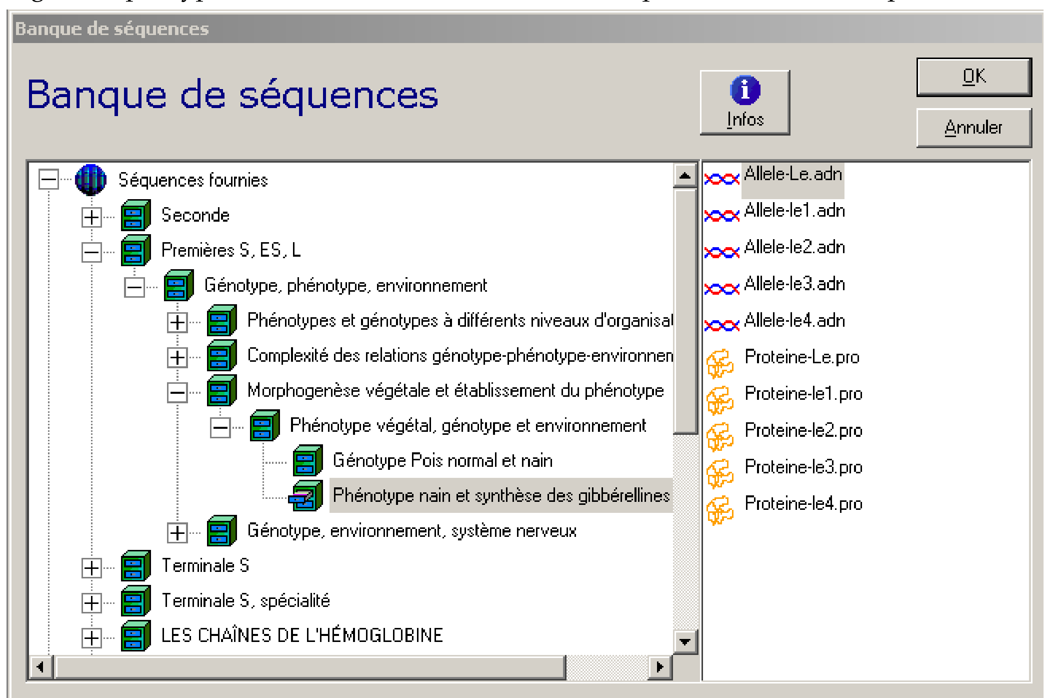
Les fenêtres des documents chargés s'empilent dans l'espace de travail d'*Anagène* mais, comme toutes les fenêtres d'*Anagène*, elles disposent d'un bouton de réduction en icône et de la possibilité de modifier leurs dimensions tout en respectant les proportions du document.

Il est aussi possible de charger des documents textuels, iconographiques (images aux formats BMP, WMF, RLE, DIB, GIF ou JPG et animations de type AVI, MOV, MPG, MP2, MPEG, ou ASF) et même des programmes exécutables propres à l'utilisateur qui pourront enrichir les possibilités d'*Anagène*. Cette fonction est assurée par la commande **Ouvrir** du menu **Fichier** qui assure aussi l'ouverture de fichiers de séquences, de résultats de traitements, de thèmes d'étude enregistrés en d'autres lieux sur différents supports informatiques (disquette, disque dur, etc.) et également de fichiers HTML. Il est à noter que la lecture des animations et vidéos numériques n'est possible que si les pilotes correspondants sont correctement installés.

Quant aux programmes externes (extension de type .exe ou .com) ainsi associés à *Anagène*, il est possible de les exécuter à la condition que la mémoire vive du micro-ordinateur soit suffisante. Si des paramètres, voire des fichiers sont à passer au programme externe, il est nécessaire de les déclarer dans la ligne de commande accessible par le menu **Édition**. Cette ligne d'édition n'apparaît que lorsque la barre de titre du programme externe révèle par sa couleur que la fenêtre est activée. Le lancement d'un programme externe, par exemple la calculatrice de *Windows* présentée dans la copie d'écran page 17, s'obtient par un clic de la souris pointée sur le bouton intitulé **Exécuter le programme externe**. Le retour dans *Anagène* s'obtient en quittant le programme externe selon la procédure qui lui est propre. Plusieurs programmes peuvent ainsi être associés à *Anagène* et ils peuvent être également répertoriés dans la banque de programmes et documents propres à l'utilisateur.

La banque de séquences

Destinée en priorité à l'enseignant, cette banque comporte toutes les séquences utilisées pour la constitution de thèmes d'étude. À chaque séquence est associé un pictogramme spécifique en relation avec sa nature (séquence nucléique de type ADN, ARN ou séquence protéique). Dans la nouvelle banque de séquences fournies, celles-ci sont organisées par niveaux et par thèmes d'étude. Les séquences de la précédente version d'*Anagène*, organisée par type de molécules, ont été maintenues après la nouvelle banque.



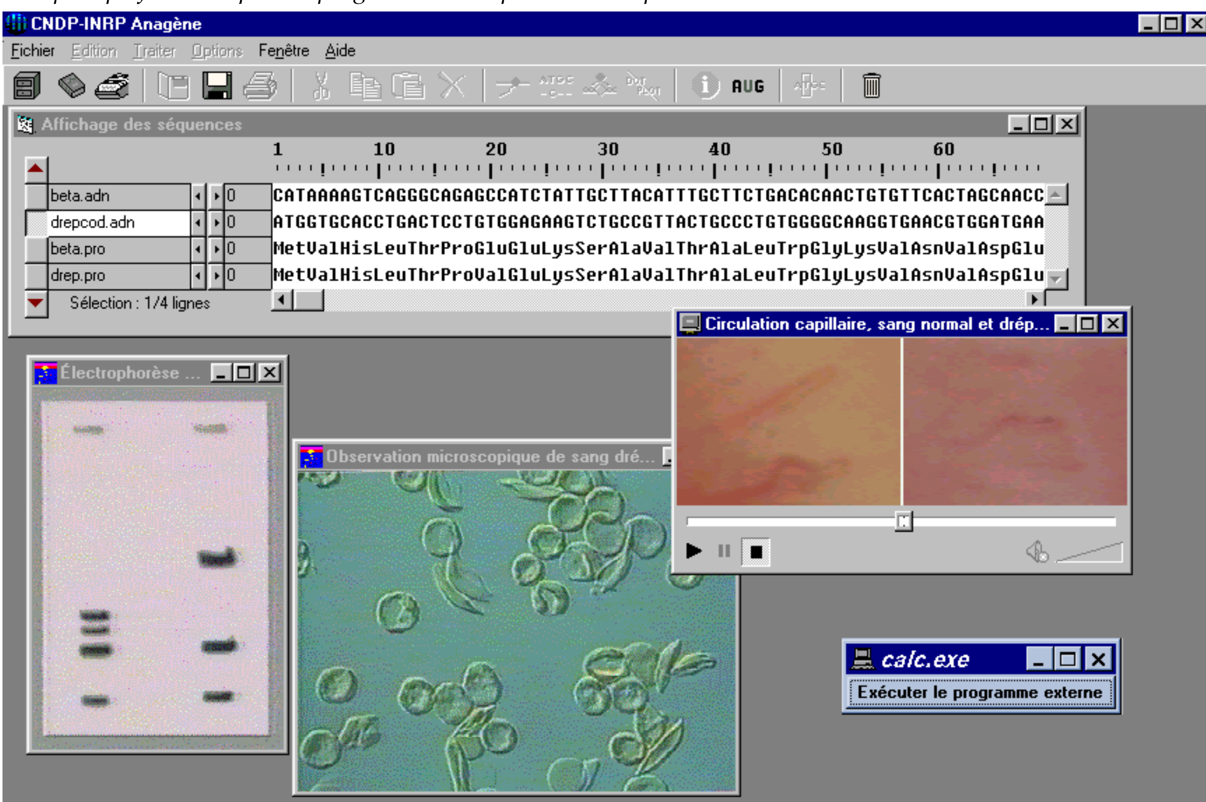
Pour charger une ou plusieurs séquences de la banque, il faut ouvrir la boîte de dialogue intitulée **Banque de séquences**. Elle est accessible par le menu **Fichier** ou en cliquant sur le premier bouton de la barre d'outils. Les séquences sont classées par dossiers, eux-mêmes ordonnés de manière arborescente. Pour les atteindre, il convient de déployer plus ou moins cette arborescence (cliquer sur les cases + ou -) et de sélectionner le dossier désiré. Dans la partie droite de la boîte de dialogue s'affiche alors la liste des séquences disponibles dans ce dossier. Une ou plusieurs séquences peuvent être sélectionnées (maintenir la touche **Ctrl** pressée), elles apparaissent alors en inversion vidéo. Un clic porté sur le bouton **Infos** renseigne sur la nature de la dernière séquence pointée qui apparaît alors encadrée. Un clic porté sur une séquence déjà sélectionnée ou la fermeture de la branche correspondante de l'arborescence la désélectionne.

La validation de la sélection d'une ou plusieurs séquences est obtenue en cliquant sur le bouton **OK** ou en appuyant sur la touche **Entrée.** Elle provoque la fermeture de la boîte de dialogue et le chargement des séquences sélectionnées dans la fenêtre d'**Affichage des séquences** qui apparaît alors à l'écran. La répétition de cette procédure permet d'ajouter de nouvelles séquences dans cette fenêtre. Le bouton **Annuler** provoque la fermeture de la boîte de dialogue et les séquences qui ont pu être sélectionnées ne sont pas chargées.

L'enseignant peut ainsi constituer des regroupements de séquences pour constituer des thèmes d'étude personnalisés.

LES DONNÉES PERSONNELLES DE L'UTILISATEUR

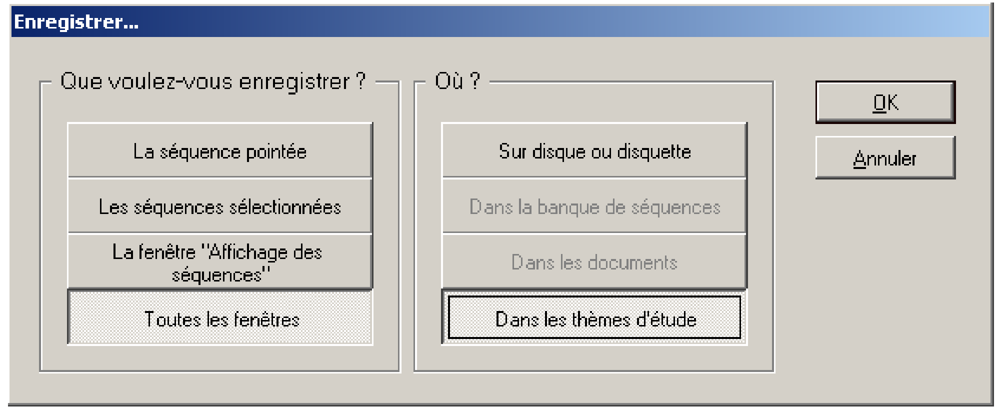
Anagène *est un logiciel ouvert qui permet d'intégrer, par le biais de menus personnalisés, les séquences créées ou chargées par l'utilisateur, les documents textuels ou iconographiques dont il dispose ainsi que les nouveaux thèmes d'étude qu'il peut souhaiter concevoir pour le public auquel il s'adresse.* Anagène *constitue ainsi un outil de travail d'autant plus performant que des programmes complémentaires peuvent lui être associés.*



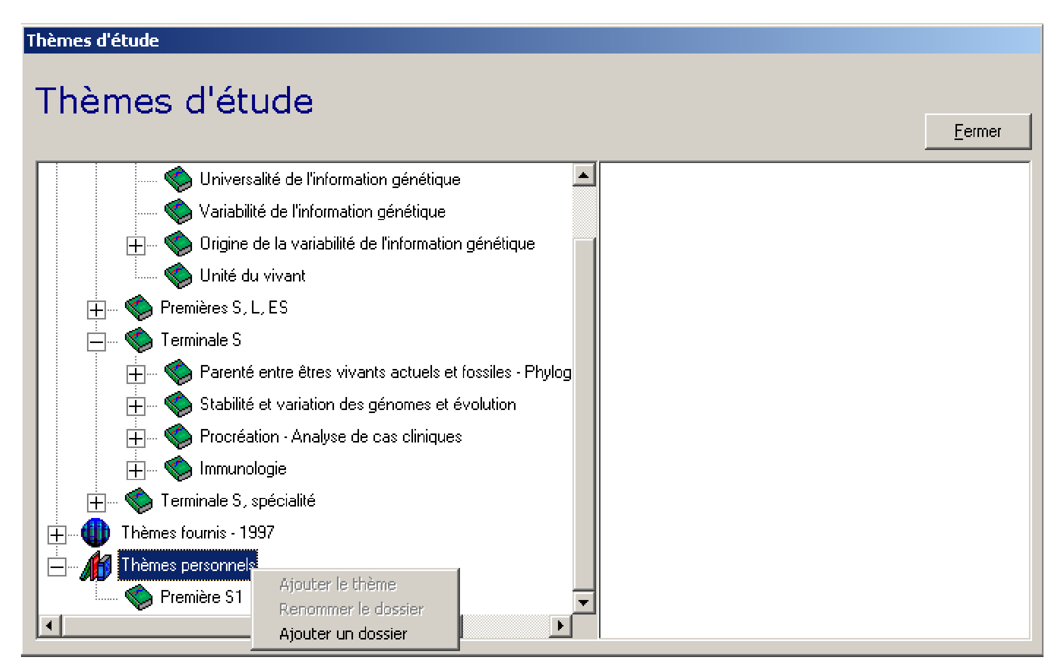
La méthode consiste à choisir dans l'espace de travail les éléments qui viendront enrichir les banques de données personnelles. Ces éléments peuvent être de différents types : séquences isolées, fenêtre de séquences, résultat d'un traitement, fenêtres avec les résultats de plusieurs traitements, documents textuels, images fixes, animations, programmes, ensemble des fenêtres présentes dans l'espace de travail.

Pour sauvegarder l'un ou l'autre de ces éléments, il suffit d'utiliser la commande **Enregistrer** accessible dans le menu **Fichier** ou plus simplement de cliquer sur le bouton disquette de la barre d'outils.

La boîte de dialogue **Enregistrer…** apparaît alors. Après avoir sélectionné ce qui doit être enregistré, il faut choisir la destination des données. Ce choix dépend de la nature des données et des objectifs.

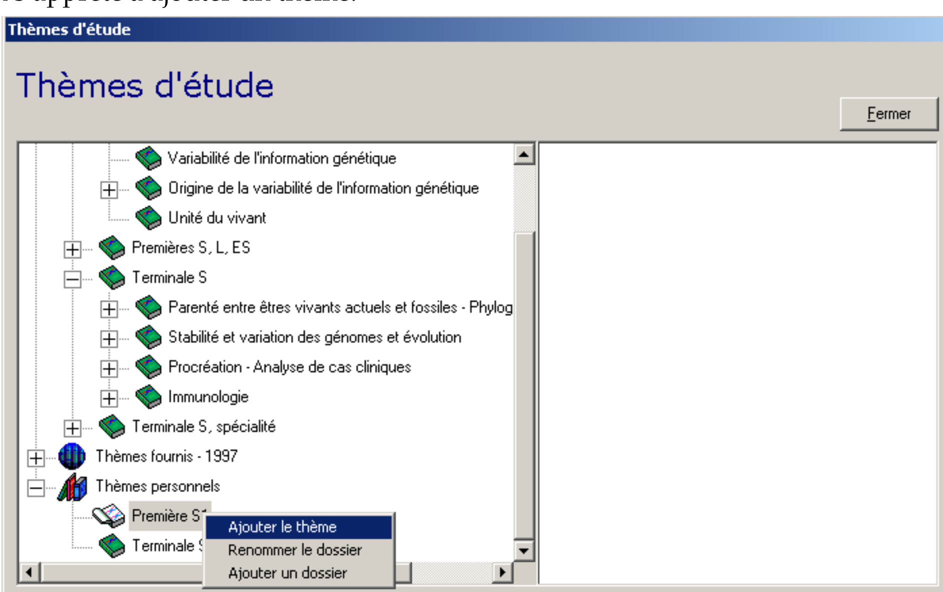


Après définition du choix de l'une des trois banques de données – un thème d'étude par exemple – et confirmation obtenue après validation, la fenêtre du menu correspondant est ouverte.



Il faut commencer par ajouter un dossier et lui donner un nom. Ensuite, on pourra ajouter un thème d'étude qui recevra aussi un intitulé.

Sur la copie d'écran suivante, on voit que deux noms de dossiers ont déjà été enregistrés. À la classe de première S1, on s'apprête à ajouter un thème.



Les libellés des thèmes et des dossiers sont modifiables. La suppression d'un dossier n'est possible que s'il ne contient aucun thème. La suppression d'un thème s'obtient en effaçant son nom dans la ligne d'édition.

L'ajout de séquences

Les séquences que l'on souhaite ajouter au menu des séquences personnelles peuvent avoir trois origines différentes.

Les séquences fournies

Ces séquences, peuvent être chargées à partir de la banque de séquences ou des thèmes d'étude fournis. Pour mettre ces séquences dans le dossier d'un utilisateur, il convient de les sélectionner dans la fenêtre d'affichage/édition d'*Anagène* et d'en demander l'enregistrement dans la banque de séquences personnelles.

Les séquences créées avec l'éditeur d'Anagène

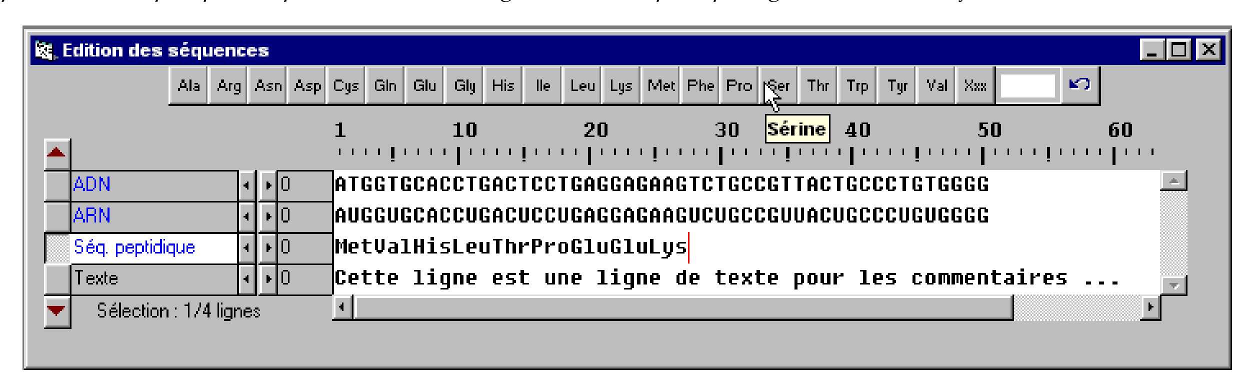
La commande **Créer** du menu **Fichier** fait passer la fenêtre d'affichage de séquences en mode édition. Au moment de la création, choisir l'un des quatre types possibles : ADN, ARN, Peptidique ou Texte sachant que l'on peut modifier le nom proposé par défaut.

La validation en cliquant sur le bouton **OK** ou en appuyant sur la touche **Entrée** déclenche l'ouverture de la fenêtre **Édition des séquences**.

Selon les cas, « l'alphabet » disponible est limité aux caractères autorisés :

- A, T, G, C, N pour les séquences d'ADN, la lettre N désignant les bases non définies ;
- A, U, G, C, N pour les séquences d'ARN, la lettre N désignant les bases non définies ;
- le code 3 lettres ou 1 lettre adopté au lancement d'*Anagène* pour désigner les acides aminés des séquences peptidiques, les lettres Xxx ou X désignant les acides aminés non définis ;
- les caractères alphabétiques, numériques et les signes de ponctuation pour les lignes de texte servant de commentaire.

En cas d'erreur de saisie lors de la création de séquences nucléiques ou peptidiques, l'appui sur le bouton situé à l'extrémité droite de « l'alphabet » efface la dernière phase de saisie. Il est donc recommandé, lors de la saisie de longues séquences, de cliquer périodiquement hors de la ligne d'écriture pour protéger les données déjà saisies.



Les séquences importées

La commande **Ouvrir** du menu **Fichier** permet de charger les séquences sauvegardées sur disque, disquette ou autre support informatique. Ces séquences peuvent avoir été créées avec un éditeur ou un traitement de texte et sauvegardées au format texte avec pour extension ADN, COD, ARN, PRO. Si *Anagène* n'a pas pu déterminer la nature d'un fichier de séquence, il propose à l'utilisateur de l'identifier et même de le modifier s'il le souhaite. Il reconnaît en particulier les séquences au format du logiciel *Seqaid*.

La suppression d'une séquence ajoutée par l'utilisateur dans la banque de séquences personnelles s'effectue en six étapes :

- afficher une séquence,
- demander son enregistrement grâce à la commande **Enregistrer** du menu **Fichier,**
- sélectionner son nom,
- demander à **Renommer la séquence** en cliquant sur le bouton droit de la Souris,
- effacer le libellé de la séquence et valider.
- accepter la suppression en cliquant sur le bouton **OK**.

La suppression d'un dossier ne se fait que si la liste des séquences personnelles de ce dossier est vide.

L'ajout de thèmes d'étude

Malgré la variété des thèmes d'étude fournis, il est possible d'en créer de nouveaux adaptés à une stratégie pédagogique personnelle ou à un public déterminé.

*Un thème d'étude est constitué par l'enregistrement de l'une ou de toutes les fenêtres présentes dans l'espace de travail d'*Anagène*, même si elles sont réduites en icônes.*

Ces fenêtres peuvent contenir des chargements effectués à partir des banques de séquences, des thèmes d'études ou des programmes et documents fournis ou personnels, des compléments en ligne téléchargés. Ces fenêtres peuvent aussi résulter des traitements effectués sur les séquences. Enfin, elles peuvent être importées grâce à la commande **Ouvrir** du menu **Fichier** qui propose les types de fichiers suivants :

- séquences nucléiques et protéiques qui seront chargées dans la fenêtre d'affichage de séquences d'*Anagène* ;
- traitements effectués et fenêtres de résultats des conversions, des comparaisons, des actions enzymatiques et des graphiques de ressemblance et sauvegardées lors de précédentes sessions de travail avec *Anagène ;*
- textes, fichiers créés avec des éditeurs ou des traitements de texte et enregistrés au format texte (extension TXT) ;
- fichiers de pages au format HTM ou HTML téléchargés sur le réseau Internet ;
- images, provenant en particulier de banques de la collection « Imagerie scientifique » éditée par le CNDP ou issues d'autres sources ; les formats reconnus sont : BMP, WMF, RLE, DIB, GIF ou JPG ;
- vidéos, au format AVI, MOV, MPG, MP2, MPEG ou ASF, qu'elles soient d'origine infographique ou obtenues par numérisation de vidéogrammes ;
- programmes, qu'il s'agisse des fichiers de commandes BAT ou PIF ou de programmes exécutables au format EXE ou COM ;
- thèmes d'étude, créés au cours d'autres sessions avec *Anagène* et enregistrés sur disque dur, disquette ou autre support informatique.

*La première étape de la création d'un nouveau thème d'étude personnel consiste donc à ne conserver dans l'espace de travail d'*Anagène *que les fenêtres souhaitées et en éliminant de la fenêtre d'affichage/édition les séquences inutiles. La seconde étape a pour but d'enregistrer le thème d'étude.*

La suppression d'un thème d'étude créé par l'utilisateur se fait de la même manière que la suppression d'une séquence.

L'ajout de programmes et de documents

Toute fenêtre présente dans l'espace de travail d'*Anagène,* quel que soit son contenu (programme exécutable, document iconographique ou textuel, séquences ou résultat de traitement de séquences), peut être ajoutée dans la liste des programmes et documents personnels. Il est à noter que c'est la fenêtre active, c'est-à-dire celle dont la barre de titre est en inversion vidéo, qui constituera, selon sa nature, le programme ou le document à enregistrer. On utilisera à cet effet la commande **Enregistrer** du menu **Fichier**.

Ici encore, ce programme ou ce document aura pour origine l'une des banques fournies avec le logiciel ou pourra être importé grâce à la commande **Ouvrir** du menu **Fichier.**

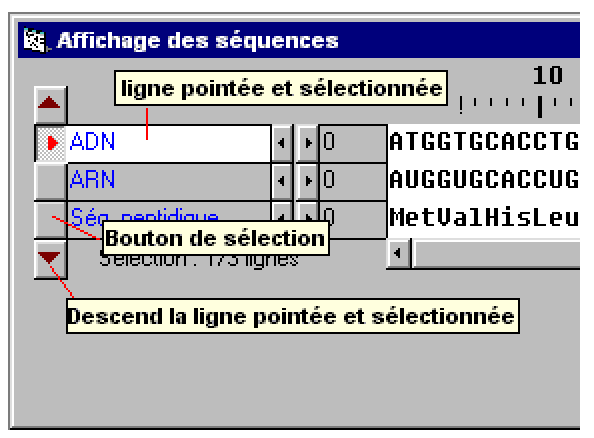
La suppression d'un programme ou d'un document ajouté par l'utilisateur s'effectue de la même manière que la suppression d'une séquence.

L'ÉDITION ET LES TRAITEMENTS DE SÉQUENCES

L'édition de séquences

Il convient de bien noter au préalable que :

- les banques de séquences et de thèmes fournies ne sont pas modifiables par l'utilisateur enseignant ou élève avec le logiciel *Anagène* lui-même ;
- au début d'une session de travail, les séquences sont protégées contre une éventuelle altération, et il convient donc de les déprotéger dans le menu **Options** si l'on veut utiliser des traitements qui modifient ces séquences ;
- la modification de séquences ne peut avoir lieu que dans la fenêtre d'édition des séquences ; le seul fait de décocher **Protéger les données** transforme la fenêtre **Affichage des séquences** en **Édition des séquences** et réciproquement ;
- les commandes d'édition sont regroupées dans le menu **Édition** et directement accessibles pour certaines à partir de leurs icônes ; noter la présence dans ce groupe de la commande d'inversion de séquences ;
- il est recommandé de dupliquer toute séquence avant de la modifier par l'usage du presse-papiers (**Copier** et **Coller**) ; les modifications étant limitées par l'« alphabet » propre à chaque type de séquence, l'espace n'étant autorisé que dans les lignes de commentaires ;
- l'ordre vertical dans lequel se trouvent les séquences dans la fenêtre **Affichage des séquences** ou **Édition des séquences** peut être modifié :



  - en triant (commande **Trier les séquences…** du menu **Édition**) les séquences sélectionnées par ordre alphabétique croissant, inverse, ou par succession de types de séquences (ADN-ARN-Peptidique ou Peptidique-ARN-ADN) ;
  - en cliquant sur les flèches situées en haut et en bas des boutons de sélection de la liste des séquences, la ligne pointée et sélectionnée pourra monter ou descendre dans cette liste de séquences ;
- le décalage d'une séquence est possible en cliquant sur les boutons de décalage.

Les traitements de séquences

- Les commandes des traitements se trouvent dans le menu **Traiter** et sont regroupées en quatre catégories : conversion de séquences, comparaison de séquences, action enzymatique et graphique de ressemblance.
- Un traitement n'est applicable que si un nombre suffisant de séquences déjà sélectionnées est atteint et si ces séquences sont de même nature ; pour sélectionner une séquence, il faut cliquer sur son bouton de sélection, ce qui provoque un affichage en inversion vidéo de son libellé.
- Après traitement, les séquences sont désélectionnées et il convient d'effectuer une nouvelle sélection pour appliquer de nouveau un traitement ; chaque traitement donne lieu à un choix d'options à valider dans une boîte de dialogue spécifique ; ces options restent mémorisées au cours d'une session de travail.
- Les résultats d'un traitement sont par défaut placés dans la ou les fenêtres spécifiques du traitement ; lorsqu'un traitement engendre de nouvelles séquences qui peuvent à leur tour être l'objet d'un traitement, celles-ci peuvent être dirigées vers la fenêtre d'**Affichage des séquences** ou d'**Édition des séquences** à la suite des séquences déjà présentes, à la condition que soit cochée la case accompagnée du message « Placer le résultat dans la fenêtre Affichage/édition ». Exemple : convertir des séquences nucléiques en séquences peptidiques et comparer ces dernières entre elles.

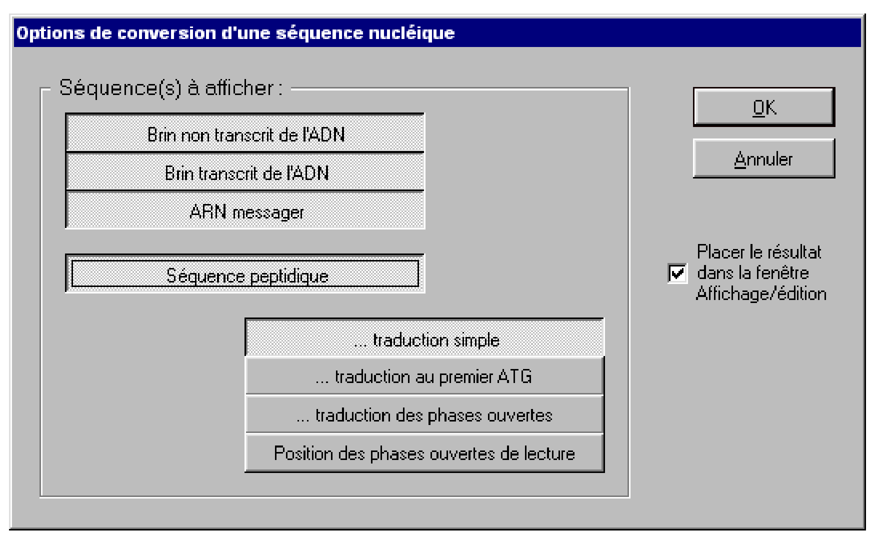
La conversion de séquences

Cette fonction peut être appliquée sur les séquences nucléiques ou peptidiques et simultanément sur plusieurs séquences si elles sont de même nature.

Si les séquences sélectionnées appartiennent à de l'ADN ou de l'ARN, une boîte de dialogue spécifique propose de cocher les conversions souhaitées et la fenêtre de destination.

Pour une conversion en séquence peptidique la traduction simple est proposée par défaut, mais on peut faire un autre choix.

Il est important de noter qu'il ne s'agit pas de « traduction biologique » dans la mesure où cette conversion s'effectue du premier triplet jusqu'au premier codon Stop rencontré en respectant le code génétique mais sans tenir compte de la présence d'un codon d'initiation. Pour que la traduction commence effectivement au codon initiation, choisir alors ***Traduction au premier ATG****.*



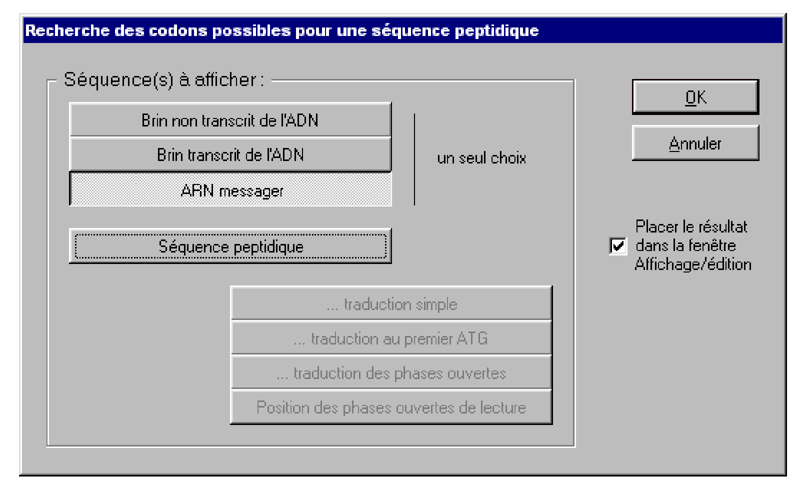
L'option **Traduction des phases ouvertes** assure les conversions correspondantes d'une séquence du premier codon d'initiation rencontré jusqu'au premier codon de terminaison, en commençant respectivement par la première base, la seconde et la troisième (les trois cadres de lecture).

L'option **Position des phases ouvertes de lecture** permet de repérer les codons spécifiques de l'initiation (symbolisés par >>>) et de la terminaison (symbolisés par ~~~).

On peut aussi, avec les commandes du menu **Édition, Copier** une sélection effectuée dans une séquence et la **Coller** dans la fenêtre d'édition en tant que nouvelle séquence à traiter.

Il faut garder à l'esprit que le logiciel Anagène *est un outil qui assure des traitements définis à l'initiative de l'utilisateur et que l'interprétation des résultats lui incombe.*

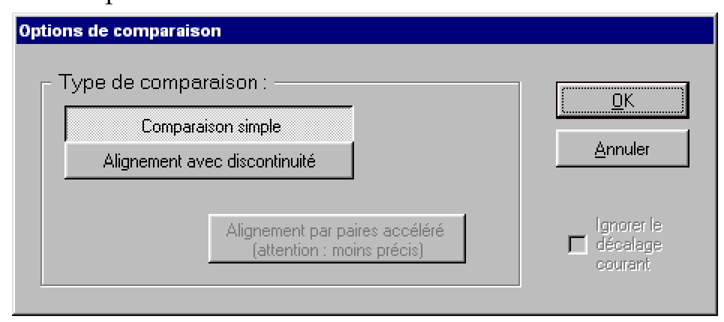
Si les séquences sélectionnées sont de nature peptidique, la seule option de conversion autorisée consiste à déterminer en ADN ou en ARN les codons possibles pour chaque acide aminé.



La comparaison de séquences

Pour être appliqué, ce traitement nécessite que deux séquences au moins, toutes les deux de nature nucléique ou peptidique, soient au préalable sélectionnées dans la fenêtre d'affichage/édition.

Deux types de comparaison sont possibles :



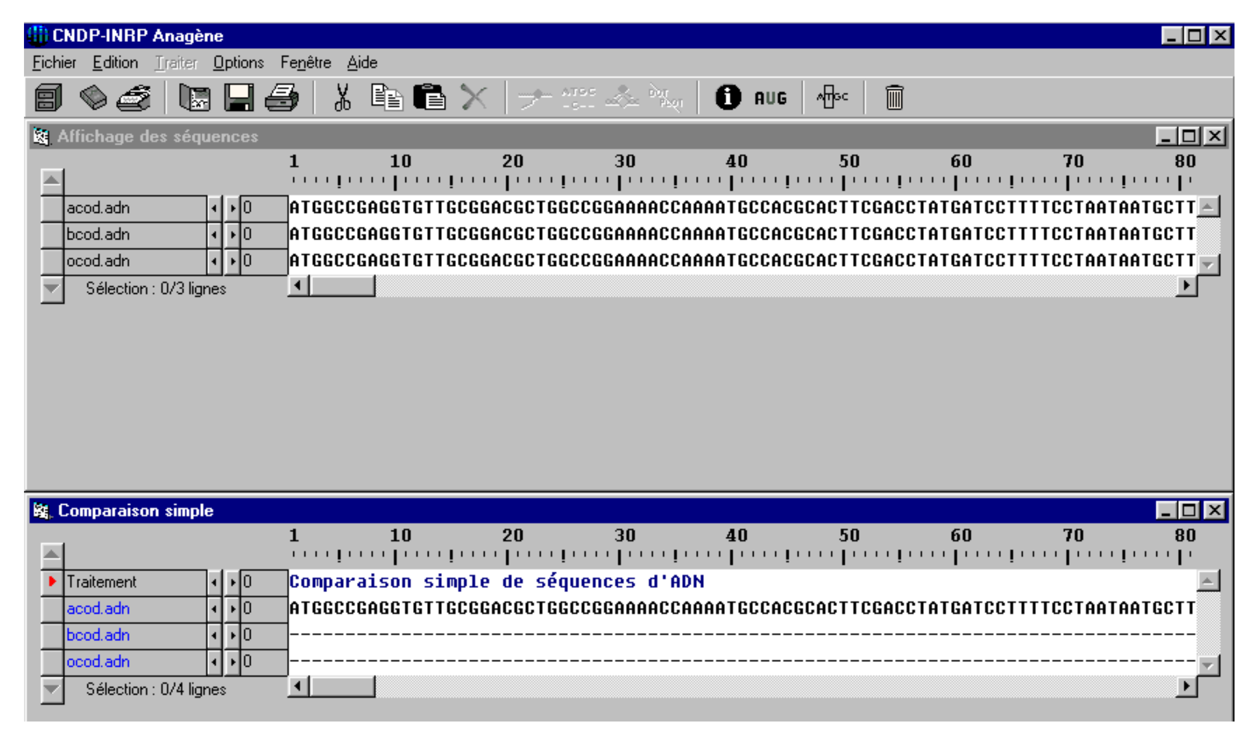
- Comparaison simple à partir de la première position dans chacune des séquences ; ce mode est plus particulièrement indiqué pour comparer, par exemple, des allèles d'un même gène comportant le même nombre de bases.

  Remarque : si un décalage des séquences a été introduit avant la comparaison, il peut être ignoré en cochant la case Ignorer le décalage courant dans la boîte de dialogue Options de comparaison.

- Alignement avec discontinuité : l'algorithme de comparaison utilisé par les chercheurs en biologie moléculaire fait intervenir plusieurs paramètres définis par défaut ; le principe consiste à arranger les séquences de manière à obtenir le plus grand nombre d'identités ; ce mode est indiqué lorsque les séquences ne sont pas de même longueur ;

  Alignement avec discontinuités par paires accéléré : l'algorithme de comparaison utilisé avec cette option est plus ancien, plus rapide, mais moins précis.

Le résultat du traitement s'affiche dans une fenêtre spécifique qui se place, comme pour les autres traitements appliqués, en dessous de la fenêtre d'affichage des séquences.

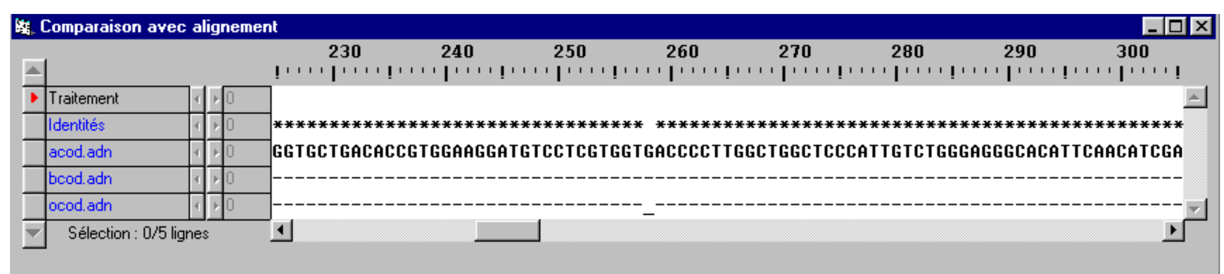


La première des séquences sert de référence pour la comparaison. L'information accessible par le bouton correspondant de la barre d'outils reconnaissable à son « Point **i** » fournit, à propos de la ligne pointée, le degré de similitude entre les séquences et la signification des symboles utilisés.

Dans le cas d'une comparaison simple, il est possible, dans la fenêtre de résultat, de provoquer le décalage d'une séquence en cliquant sur les boutons de décalage et d'obtenir immédiatement les modifications du résultat de la comparaison à la séquence de référence.

L'alignement avec discontinuités nécessite des temps de calcul d'autant plus importants que les séquences sont nombreuses, longues et différentes.

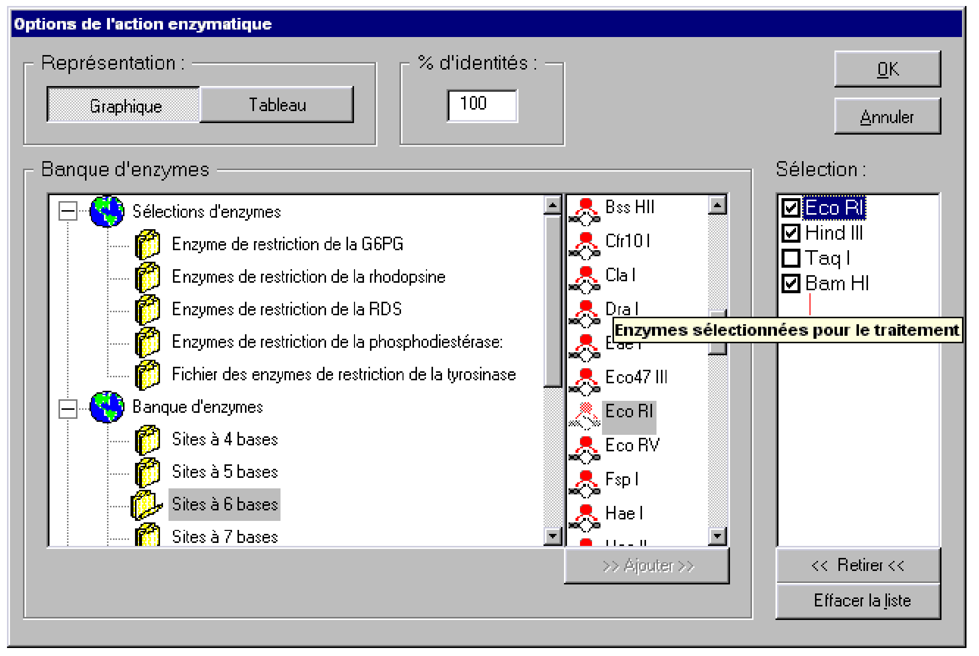
Il convient d'explorer les séquences avec la barre de défilement horizontal pour découvrir les régions plus ou moins semblables.



Remarque : si l'on souhaite, pour une comparaison, changer de séquence de référence, il suffit, avec les flèches situées en haut et en bas des boutons de sélection de la fenêtre d'affichage des séquences, de remonter ou de descendre la séquence pointée et sélectionnée. On a intérêt à bien choisir l'ordre des séquences en fonction du degré de ressemblance entre elles. Si la comparaison porte sur trois séquences et que deux d'entre elles sont semblables, leur ressemblance sera plus évidente si l'une des deux sert de référence.

L'action enzymatique

L'analyse avec des enzymes de restriction s'applique aux séquences d'ADN sélectionnées au préalable dans la fenêtre d'affichage/édition. Le choix des enzymes s'effectue dans la boîte de dialogue suivante.



Cette boîte de dialogue se présente comme une série de sélections préétablies d'enzymes de restriction utiles au traitement de certaines séquences dans le cadre d'études de prévisions en génétique humaine et d'une banque d'enzymes organisée en dossiers dans lesquels les enzymes se trouvent classées en fonction de la longueur du site de coupure qu'elles reconnaissent. Quand un dossier est ouvert, les enzymes sont à sélectionner dans la liste qui s'affiche au milieu et à **Ajouter** dans la fenêtre de droite.

Dans cette fenêtre, à droite, avec **Sélection** d'enzymes, il est possible :

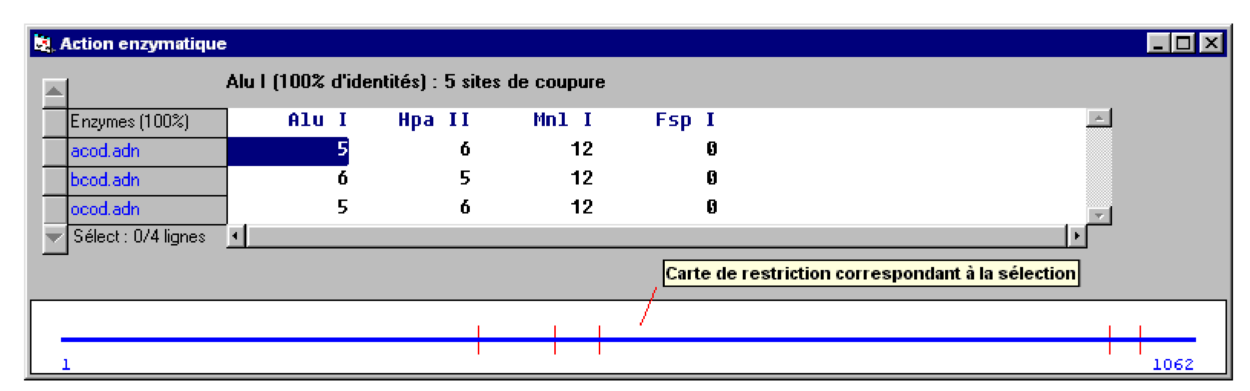
- d'ajouter de nouvelles enzymes présentes dans d'autres dossiers ;
- de décocher chacune d'elles en cliquant sur la coche ;
- de retirer l'enzyme sélectionnée en cliquant sur le bouton **Retirer** ;
- de retirer la totalité des enzymes de cette liste en cliquant sur le bouton **Effacer la liste**.

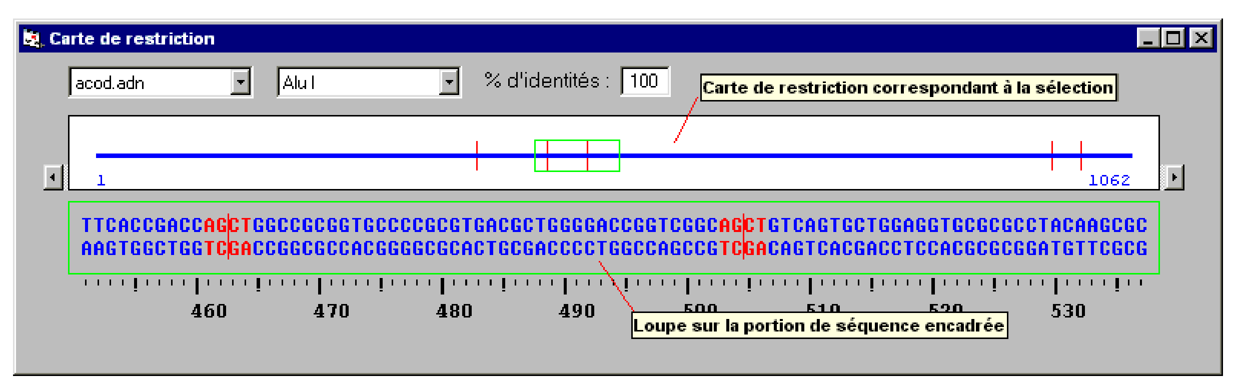
De plus, cette sélection peut être mémorisée en cliquant sur le bouton droit de la Souris au niveau de **Lots personnels** (situé tout en bas de l'arborescence), puis en cliquant sur **Ajouter un lot personnel** et en lui donnant un nom. Ainsi ces fichiers d'enzymes pourront être ouverts lors d'une nouvelle session de travail en cliquant sur le bouton **Fichier**.

Pour supprimer un fichier d'enzymes qui a été importé dans les lots personnels (par la fonction **Importer**) ou bien créé à partir des enzymes disponibles, il faut, après y avoir accédé par la commande **Action enzymatique** du menu **Traiter**, utiliser la fonction **Renommer le lot** du menu contextuel. Cette fonction apparaît quand on clique du bouton droit alors que le pointeur de la Souris est sur le nom du lot d'enzymes. Il faut alors effacer totalement le nom pour faire apparaître la boîte de dialogue **Supprimer** et confirmer la suppression. Il peut arriver que cette méthode ne fonctionne pas toujours à souhait. Dans ce cas, il faut supprimer le ou les fichiers d'enzymes qui se trouvent dans le dossier **Sauve**. Lors du lancement d'*Anagène*, ces enzymes ne sont plus alors référencés dans les lots personnels.

Deux modes de représentation de l'action enzymatique sont prévus : la visualisation sous la forme d'un tableau ou d'un graphique, les deux modes pouvant être retenus simultanément.

Après validation de la liste des enzymes sélectionnées, les fenêtres de résultat se superposent en dessous de la fenêtre d'affichage/édition des séquences.





On peut voir les deux fenêtres avec les commandes du menu **Fenêtre** d'*Anagène* qui permettent divers affichages dont le mode **Mosaïque.** Il est à noter qu'en présence de fenêtres contenant des documents iconographiques, ce mode ne peut être parfait, les proportions des images étant toujours respectées.

La fenêtre **Action enzymatique** présente le mode d'affichage **Tableau** et montre le nombre de sites reconnus par chaque enzyme de la sélection active (enzymes cochées dans la liste **Sélection d'enzymes**) agissant sur chaque séquence nucléique sélectionnée.

La fenêtre **Carte de restriction** présente le mode d'affichage **Graphique** et illustre l'action d'une enzyme de restriction sur une des séquences traitées. On peut faire varier soit la séquence choisie, soit l'enzyme. Il faut explorer la carte de restriction en déplaçant la « loupe » le long de la séquence pour y découvrir les sites de coupure.

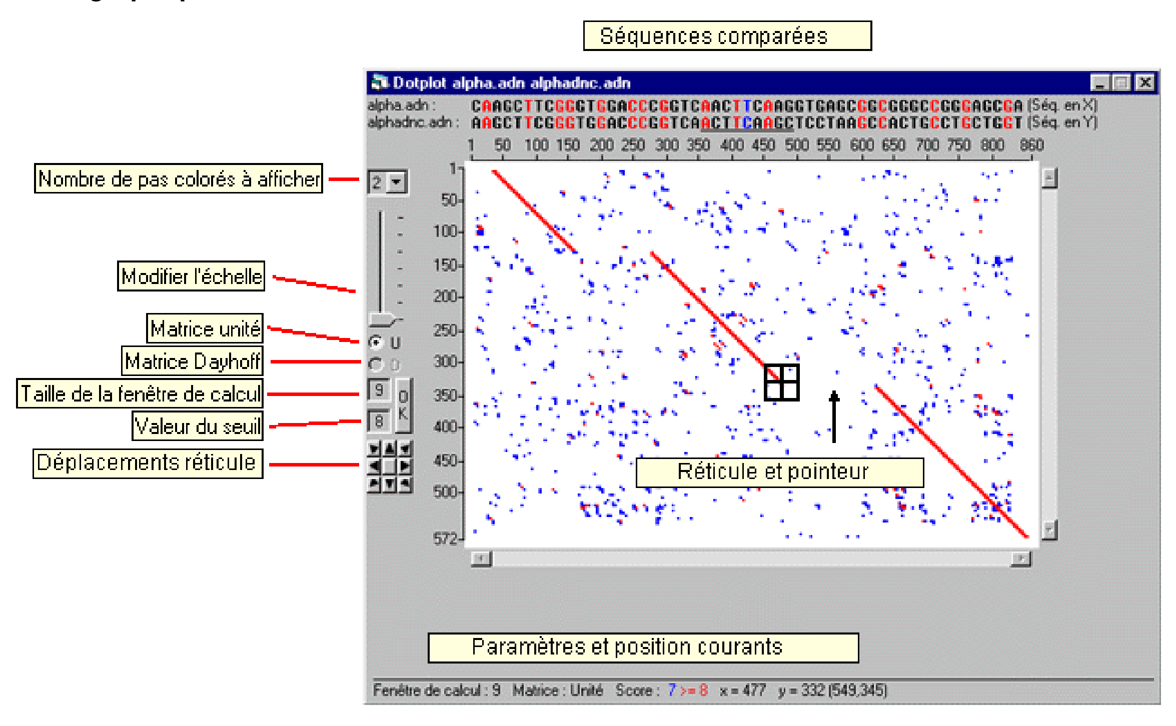
La valeur du **% d'identités** permet de rechercher des sites plus ou moins strictement identiques à celui que chaque enzyme détecte.

Bien d'autres données associées à ces tableaux de résultats et accessibles par le bouton **Informations sur la ligne pointée…** sont disponibles tant sur les sites d'action que sur les fragments de séquences d'ADN obtenus. Comme pour les autres fenêtres d'information, il est possible d'en imprimer le contenu ou de le copier dans le presse-papiers de *Windows*.

Le graphique de ressemblance

Le graphique de ressemblance (appelé aussi dotplot dans le domaine de la bio-informatique) est une représentation visuelle en forme de matrice de points permettant de comparer deux séquences pour détecter les régions ayant une forte similarité. Les deux séquences sont placées le long des axes du graphique. On peut aussi comparer une séquence avec elle-même (sa copie), ce qui donne les répétitions internes à la séquence.

Création du graphique



Des deux séquences sélectionnées, celle située au-dessus de l'autre dans la fenêtre Affichage des séquences déterminera les valeurs de l'axe des abscisses (séquence en X), et l'autre, les valeurs de l'axe des ordonnées (séquence en Y).

Pour générer ce graphique, il est nécessaire de sélectionner deux séquences et pas davantage, dans la fenêtre affichage des séquences. Les deux séquences doivent être de même type, soit nucléiques, soit protéiques.

Cliquer ensuite sur le bouton **DotPlot** de la barre d'icônes ou sur la commande **Graphique de ressemblance** du menu **Traiter** : la fenêtre du graphique de ressemblance s'affiche alors.

Chaque point de couleur dans le graphique de ressemblance signifie qu'à l'intersection de la ligne et de la colonne qui déterminent les coordonnées de ce point, une identité de nucléotide ou d'acide aminé existe dans les deux séquences.

Une suite de points disposés en diagonale indique les régions de similarité entre les deux séquences.

Au-dessus du graphique, une portion colorée des deux séquences est affichée. Cette portion est déterminée par la position courante du réticule.



La couleur rouge est affectée aux symboles superposés identiques d'une séquence à l'autre, la couleur noire aux éléments différents et la couleur bleue à la position courante du réticule (les éléments peuvent y être identiques ou différents).

La longueur du trait horizontal situé sous les séquences correspond à la taille de la fenêtre de calcul.

Dans le cas de séquences protéiques, c'est le code 1 lettre qui est utilisé pour désigner les différents acides aminés.

Il est possible de modifier la position du réticule :

- en cliquant directement dans la matrice ;
- en utilisant les flèches de direction du clavier ;
- en cliquant sur les boutons de déplacement.

Il est possible de déplacer la matrice dans la fenêtre de visualisation :

- en utilisant les barres de déplacement situées à droite et au-dessous de la fenêtre de visualisation ;
- en cliquant sur les boutons de déplacement après avoir cliqué sur le bouton central (il s'affiche maintenant en noir).

Il est possible de faire varier la portion de matrice affichée dans la fenêtre de visualisation en modifiant l'échelle par déplacement du curseur.

Optimiser la recherche des similarités

Diverses actions de filtration permettent de modifier notablement l'affichage du graphique pour renforcer la mise en valeur des similarités en diminuant le bruit de fond.

En modifiant la taille de la fenêtre de calcul, la valeur du seuil et le nombre de pas colorés à afficher, il est possible d'appliquer un nouveau filtre d'affichage aux séquences.

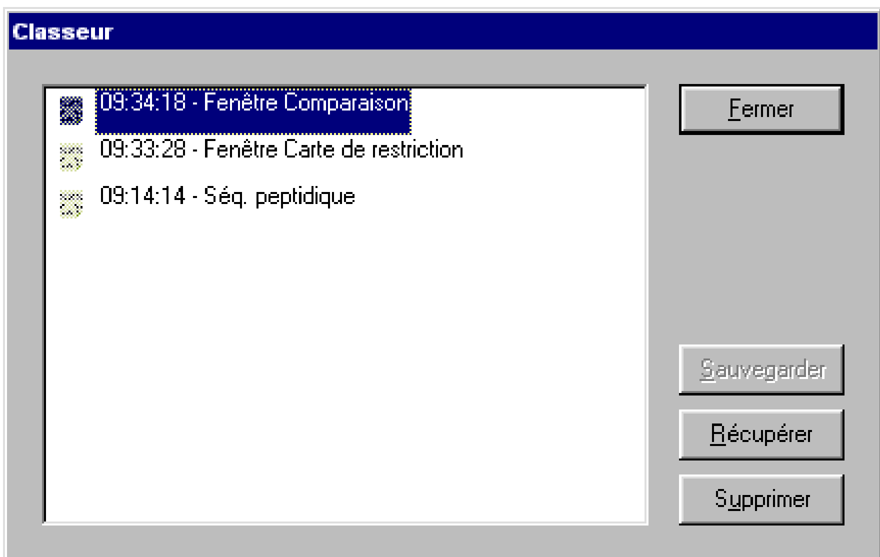
Sur une portion des deux séquences de longueur égale à la taille de la fenêtre de calcul, un score est calculé ; il vaut la somme du nombre de couples présentant une identité de symboles (nucléotides ou acides aminés) écrits en couleur rouge augmentée éventuellement de un si le couple de couleur bleue présente une identité.

Si ce score est égal ou supérieur à la valeur seuil, un point rouge sera placé dans le graphique de ressemblance. Si le nombre de pas colorés est supérieur à un, les scores calculés, immédiatement inférieurs à la valeur seuil, produiront l'affichage sur le graphique de ressemblance d'un point de couleur bleue, verte ou jaune. Dans le seul cas d'une comparaison de séquences protéiques, il est possible, à la place de la matrice unité, d'appliquer la matrice de substitution de Dayhoff qui représente assez bien les probabilités de substitution d'un acide aminé par un autre lors de l'évolution des protéines.

LE CLASSEUR D'*ANAGÈNE*

Il s'agit d'un utilitaire qui permet à l'élève de sauvegarder simplement et *temporairement* les séquences qui, trop nombreuses, viennent à encombrer la fenêtre d'affichage des séquences ou les résultats des traitements effectués.

Sa simplicité tient au fait qu'il n'y a pas à parcourir l'arborescence du disque dur pour y enregistrer des fichiers. Après avoir sélectionné les séquences de la fenêtre d'affichage ou la fenêtre de résultat à sauvegarder, il suffit de cliquer sur le bouton correspondant de la barre d'outils pour l'ouvrir. Les données pourront alors être intégrées au classeur par l'ajout de leur libellé et de l'heure correspondante.



En fin de session, ce classeur est vidé après affichage d'un message d'avertissement invitant à en visiter le contenu pour récupérer si besoin des données. Ainsi, le disque dur n'est pas surchargé avec des fichiers inutiles.

En cas de coupure inopinée de l'électricité ou de réinitialisation du micro-ordinateur, le contenu du classeur est récupérable dès le lancement d'*Anagène*.

GUIDE DE RÉFÉRENCE

LES COMMANDES DES MENUS

Menu Fichier

Banque de séquences affiche sous forme arborescente la liste des séquences disponibles.
Thèmes d'étude affiche sous forme arborescente la liste des thèmes d'étude prédéfinis.
Programmes et documents affiche sous forme arborescente la liste correspondante.
Créer permet de créer une séquence nucléique, protéique ou une ligne de texte.
Ouvrir permet de choisir un fichier dans l'arborescence des répertoires et de l'ouvrir dans l'espace de la fenêtre-mère d'*Anagène*.
Importer permet d'entrer les données chargées directement dans la banque.
Enregistrer assure la sauvegarde de séquences, d'une fenêtre ou de l'ensemble.
Imprimer imprime les séquences choisies et les résultats de traitements appliqués.
Quitter permet, après confirmation, de quitter définitivement le logiciel *Anagène*.

NB : Il est toujours proposé d'enregistrer une fenêtre dont le contenu a changé avant de la fermer.

Menu Édition

Couper permet de mémoriser une sélection dans le presse-papiers ; de plus, il efface la sélection.

Copier permet simplement de mémoriser une sélection dans le presse-papiers.

Coller permet de recopier à l'emplacement du curseur, la sélection mémorisée dans le presse-papiers.

Effacer permet de supprimer définitivement une sélection.

Sélectionner tout assure la sélection de l'ensemble des séquences d'une fenêtre.

Désélectionner tout annule les sélections effectuées dans une fenêtre.

Rechercher permet de rechercher, dans la séquence où se situe le curseur, les occurrences d'une chaîne de caractères.

Minuscules convertit en minuscules la sélection effectuée dans la fenêtre d'édition des séquences.

Majuscules convertit en majuscules la sélection effectuée dans la fenêtre d'édition des séquences.

Inverser la séquence provoque l'inversion de la séquence pointée dans la fenêtre d'édition des séquences.

Trier les séquences organise dans l'ordre choisi les séquences de la fenêtre d'affichage/édition des séquences.

Menu Traiter

Convertir les séquences affiche ADN brin non transcrit, ADN brin transcrit, ARNm, séquence peptidique, codons possibles pour un polypeptide.

Comparer les séquences compare des séquences à une séquence de référence à partir de la première position et réalise l'alignement avec discontinuités.

Action enzymatique révèle les sites d'action d'enzymes de restriction sur des séquences d'ADN.

Graphique de ressemblance compare deux séquences en produisant une matrice rectangulaire.

Menu Options

√***Protéger les données*** empêche toute modification de séquence dans la fenêtre d'affichage.

√ ***Grand curseur*** allonge le curseur sur la hauteur de l'ensemble des séquences jusqu'à l'échelle et sur la largeur de la sélection en cours.

√***Règle en triplets*** affiche la règle graduée en triplets plutôt qu'en caractères.

√ ***Dossier par défaut*** permet de définir le chemin du dossier pour les sauvegardes temporaires.

√***Bulles d'aide*** affiche les bulles d'aide associées aux objets de l'interface.

Menu Fenêtre

Mosaïque place les fenêtres côte à côte tout en respectant les proportions des documents iconographiques.

Cascade organise les fenêtres en les superposant légèrement décalées.

Réorganiser les icônes aligne les fenêtres réduites en icônes grâce au bouton de réduction situé à droite de la barre de titre.

Les titres des fenêtres permettent d'accéder directement aux fenêtres désirées.

Menu Aide

***À propos d'*Anagène** affiche les informations relatives à l'édition du logiciel.

Rechercher de l'aide assure la recherche sur les mots clés indexés de l'aide.

Sommaire de l'aide permet d'accéder au sommaire de l'aide en ligne.

@ Complément en ligne affiche une page HTML si un navigateur récent est fonctionnel sur le micro-ordinateur. Cette page comporte des liens vers les sites du CNDP et de l'INRP ainsi qu'un lien vers un site dédié à *Anagène*. En l'absence de navigateur récent, le micro-ordinateur affiche une page comportant les adresses de ces sites.

LES ICÔNES DE LA BARRE D'OUTILS

Première série : banques de données

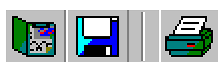


1- Afficher l'arborescence de la banque de séquences.

2- Afficher l'arborescence de la liste des thèmes d'étude.

3- Afficher l'arborescence de la liste des programmes et documents.

Deuxième série : outils de sauvegarde

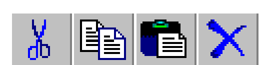


1- Voir le classeur.

2- Enregistrer les séquences sélectionnées, la fenêtre active ou l'ensemble des fenêtres.

3- Imprimer les séquences choisies et les résultats des traitements appliqués.

Troisième série : outils d'édition



1- Couper une sélection et la mémoriser dans le presse-papiers.

2- Copier une sélection et la mémoriser dans le presse-papiers.

3- Coller, à l'emplacement du curseur, la sélection mémorisée dans le presse-papiers.

4- Effacer définitivement une sélection.

Quatrième série : outils de traitement



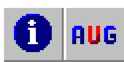
1- Convertir les séquences : affiche ADN brin non transcrit, ADN brin transcrit, ARNm, séquence peptidique, codons possibles pour un polypeptide.

2- Comparer les séquences à une séquence de référence à partir de la première position et réaliser l'alignement avec discontinuités.

3- Révéler les sites d'action d'enzymes de restriction sur des séquences d'ADN.

4- Afficher le graphique de ressemblance des deux séquences sélectionnées.

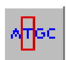
Cinquième série : accès aux informations



1- Renseigner sur la séquence pointée ou sur le résultat d'un traitement pointé.

2- Afficher le tableau interactif du code génétique en ARN.

Avant-dernière icône : gestion du curseur



Allonger le curseur sur la hauteur de l'ensemble des séquences et de la règle, et sur la largeur de la sélection en cours, ou rétablir le curseur par défaut.

Dernière icône



Après confirmation, fermer toutes les fenêtres.

LES RACCOURCIS CLAVIER

Alt F4 permet, après confirmation, de fermer toutes les fenêtres puis de quitter *Anagène.*

Ctrl F4 provoque la fermeture de la fenêtre active.

Ctrl C permet de copier une sélection et de la mémoriser dans le presse-papiers.

Ctrl X permet de couper une sélection et de la mémoriser dans le presse-papiers.

Ctrl V permet de coller à l'emplacement du curseur, la sélection mémorisée dans le presse-papiers.

Ctrl A permet de sélectionner toutes les séquences de la fenêtre active.

Ctrl B supprime ou rétablit les bulles d'aide**.**

Ctrl R permet de rechercher dans la séquence une suite de caractères préalablement sélectionnée ou saisie au clavier.

Ctrl I permet d'imprimer les séquences choisies et les résultats de traitements appliqués.

F1 affiche l'aide en ligne du logiciel.

F3 permet d'atteindre l'occurrence suivante dans une recherche sur une suite de caractères.

F4 permet d'activer le graphique de ressemblance.

F5 modifie l'aspect du curseur.

F6 modifie l'aspect de la règle graduée.

F7 assure les traitements de conversion de séquences.

F8 assure les traitements de comparaison de séquences.

F9 assure les traitements enzymatiques sur les séquences d'ADN.

F12 permet d'enregistrer les séquences sélectionnées, la fenêtre active ou l'ensemble des fenêtres.

Maj F4 organise les fenêtres en cascade.

Maj F5 organise les fenêtres en mosaïque tout en respectant les proportions des documents iconographiques.

Suppr efface définitivement une sélection.

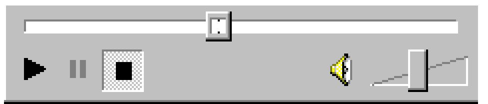
Entrée provoque le même effet qu'un clic sur le bouton ***OK*** ou sur tout autre bouton de validation de choix.

Echap provoque le même effet qu'un clic sur le bouton ***Annuler*** pour abandonner un choix.

Alt Lettre soulignée équivaut à un clic sur la commande correspondante d'un menu déroulant.

LE VISIONNEUR D'ANIMATIONS

Le visionneur d'animations permet d'agir sur le déroulement de l'animation ou de la vidéo chargée.



Lancement de la lecture de l'animation ou de la vidéo.

Pause dans la lecture de l'animation ou de la vidéo.

Arrêt de la lecture et repositionnement au début de l'animation ou de la vidéo.

Positionnement par déplacement du curseur de la lecture à un moment précis.

Activation ou désactivation du son qui accompagne l'animation ou la vidéo.

Réglage du volume sonore par déplacement du curseur le long de la barre.

SUGGESTIONS PÉDAGOGIQUES : CLASSE DE SECONDE

INTRODUCTION : LA BANQUE DE DONNÉES ET LES NOTIONS DU PROGRAMME

Les exemples supports

L'objectif de cette partie du programme est de conforter la notion d'évolution et de sensibiliser à celles d'ancêtre commun et d'origine commune de tous les êtres vivants (ces notions, abordées au collège, peuvent être renforcées en seconde : ainsi, l'étude des plans d'organisation renforcera-t-elle par exemple la notion d'ancêtre commun).

En classe de seconde, la recherche de ce qui est commun aux Vertébrés et à l'ensemble des êtres vivants amène à préciser les notions de cellule et d'information génétique, en découvrant notamment la structure et le rôle de l'ADN. Les notions de gènes et d'allèles sont ainsi renforcées.

La notion d'information génétique portée par les chromosomes a été introduite au collège. L'étude, en classe de seconde, d'une ou plusieurs expériences de transgenèse permet de conclure quant au fait que c'est l'ADN qui est le support de l'information génétique, ce qui motive l'étude de sa structure et de ses caractéristiques. La découverte de la structure de l'ADN amène à émettre l'hypothèse qu'une séquence de nucléotides peut être porteuse de l'information génétique. Les caractéristiques de l'ADN peuvent être découvertes par un travail sur la visualisation 3D de diverses molécules d'ADN (utilisation du logiciel *RasTop*), complété par la comparaison des séquences nucléiques correspondantes.

À partir de là, on peut tester l'hypothèse émise précédemment (l'information génétique réside dans la séquence de nucléotides de l'ADN) en constatant qu'il existe une relation entre la variabilité phénotypique et la variabilité de l'ADN.

Cette étude de la variabilité génétique est l'occasion de renforcer la notion de gène et d'allèles. La comparaison de deux gènes différents (par exemple cdc2 et ade2) peut alors être intéressante pour faire prendre conscience de ce qui fait la différence entre deux gènes et deux allèles d'un même gène. Il s'agit de leur localisation chromosomique, d'une part, et du taux de similitude entre les séquences — très élevé pour deux allèles, beaucoup plus faible pour deux gènes — d'autre part.

Pour chacun des exemples proposés, la démarche peut être la même :

- découvrir la variabilité phénotypique ;
- faire l'hypothèse d'une détermination de cette variabilité phénotypique par une variabilité de l'ADN ;
- comparer les séquences alléliques disponibles de façon à tester l'hypothèse émise.

Deux des exemples retenus permettent une approche expérimentale pour découvrir la variabilité phénotypique :

- gène ade2 (dont l'allèle normal s'exprime par une colonie blanche de Levures, et l'allèle muté retenu ici par une colonie rouge) ;
- gène cdc2 (l'allèle normal détermine une division cellulaire normale, un des allèles mutés retenus ici est responsable de la division précoce de la cellule et l'autre allèle muté retenu est responsable d'une division cellulaire dépendane de la température).

Le troisième exemple, celui de la drépanocytose, ne permet pas une approche expérimentale, mais il a été retenu car il permet de relier une différence phénotypique à une différence entre les allèles du gène en jeu chez un organisme pluricellulaire. De plus, cet exemple sera réinvesti et complété en classe de première lors de l'étude des relations génotype/phénotype/environnement.

En classe de seconde, le niveau du phénotype moléculaire, c'est-à-dire polypeptidique, n'est pas au programme.

Une fois découverte la variabilité de l'information génétique, se pose le problème de **l'origine de cette variabilité**. Il s'agit alors d'introduire la notion de mutation, et de distinguer ensuite mutations somatiques et mutations germinales.

La distinction entre mutations somatiques et mutations germinales peut être introduite à partir de l'exemple des cancers. Un exemple est proposé : l'implication des mutations du gène p53 dans certains cancers (certains

cancers semblent impliquer une mutation germinale du gène p53, comme c'est le cas pour le syndrome de Li-Fraumeni ; d'autres cancers semblent impliquer une mutation somatique du gène p53). Il convient bien sûr d'être très prudent lorsque cet exemple concernant les cancers est utilisé, et de bien préciser le caractère plurifactoriel de ces cancers. L'ouverture possible vers l'intérêt d'un dépistage des mutations germinales, quand il sera possible, peut amener à évoquer la médecine prédictive et préventive, même si cela n'est pas au programme de cette classe.

Pour terminer, afin de dégager l'unité du vivant, on fera prendre conscience à l'élève que des gènes importants pour les organismes sont partagés par un grand nombre d'êtres vivants, ce qui témoigne d'une origine commune :

- gène cdc2, présent chez toutes les cellules eucaryotes, qui intervient dans le contrôle du cycle cellulaire ;
- gènes homéotiques, présents chez de nombreux animaux, qui contrôlent la mise en place du plan d'organisation.

Utilisation d'*Anagène*

L'utilisation du logiciel *Anagène* en classe de seconde dans les démarches proposées peut paraître limitée. Cependant, cette utilisation, pour des séquences courtes incluses dans une séance de TP pouvant se répéter dans plusieurs séances consécutives ou non, présente l'intérêt de permettre aux élèves de prendre en main cet outil et de se familiariser avec certaines de ses fonctionnalités.

Cette initiation favorisera grandement l'utilisation de ce logiciel dans la suite de la scolarité (première S, L ou ES, terminale S), quand il s'avérera un outil essentiel utilisable pour traiter nombre de thèmes (relations génotype/phénotype/environnement, immunologie, évolution, etc.).

Enfin, il faut souligner la complémentarité des logiciels *Anagène* (traitement de données moléculaires) et *RasTop* (visualisation 3D de molécules) dans les démarches proposées, complémentarité qui est abordée ici, notamment dans le cadre de l'étude de l'universalité de la molécule d'ADN. Cette complémentarité se retrouvera dans nombre de démarches proposées en classe de première et en classe terminale.

PARENTÉ ET DIVERSITÉ DES ORGANISMES – UNIVERSALITÉ ET VARIABILITÉ DE L'INFORMATION GÉNÉTIQUE

Universalité de l'information génétique

Structure de l'ADN

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Universalité de l'information génétique** permet d'atteindre **Structure de l'ADN** qui charge le fichier adn.edi affichant six séquences nucléiques différentes, courtes, provenant de divers êtres vivants : le Rat (ADNrat), l'Homme (ADNhumain1 et ADNhumain2), la Levure (ADNLevure), la bactérie *Escherichia Coli* (ADNbacterieEC), le virus de l'herpès (ADNvirusherpes).

Fichiers des molécules en 3D

Dans le dossier 3D sous Anagene2, le répertoire **StructureADN** contient les fichiers .pdb suivants : ADNrat.pdb, ADNhumain1.pdb et ADNhumain2.pdb, ADNLevure.pdb, ADNbacterieEC.pdb, ADNvirusherpes.pdb.

Pistes d'exploitation pédagogique des séquences

Anagène n'est pas bien sûr l'outil premier pour découvrir la structure de l'ADN. Celle-ci peut-être établie avec le logiciel *RasTop*.

La visualisation 3D avec ce logiciel de différents fragments d'ADN d'origine diverse permet de dégager quelques caractéristiques de la molécule d'ADN (il est important de préciser aux élèves qu'il ne s'agit ici que de fragments, de longueurs différentes, de molécules d'ADN beaucoup plus longues en réalité) et l'universalité de sa structure (puisque l'on dispose de fragments d'ADN appartenant à des espèces variées).

La banque d'*Anagène* fournit pour chaque fragment visualisable avec *RasTop* la séquence d'un des deux brins. Cela permet d'aborder avec les élèves la façon dont l'ADN est représenté dans *Anagène*, ce qui est essentiel pour l'utilisation ultérieure des banques de séquences. Pour bien faire saisir en quoi la représentation d'un seul brin d'ADN est suffisante, on peut demander aux élèves de schématiser le fragment d'ADN complet à partir de la séquence fournie par *Anagène*. Bien entendu, puisque l'expression des gènes n'est pas au programme, et que les fragments fournis dans *RasTop* ne sont peut-être pas codants, on ne peut pas parler de brin transcrit et non transcrit en seconde.

Ces fragments d'ADN ne sont pas homologues. On peut rechercher ce qu'ils ont en commun et en quoi ils diffèrent. Une comparaison simple permet de discuter de ce qui peut être le support de l'information génétique : proportion des différents nucléotides et ordre de leur agencement (ce que l'on appelle la séquence). Classiquement, on ne retient que le second point mais, à ce stade, les deux hypothèses peuvent être retenues.

La comparaison simple de deux fragments de même longueur (ADN-humain2 et ADN-virusHerpes puisés dans le fichier adn.edi), dont la composition en bases n'est pas très éloignée, aboutit à une similitude de 22,8 %. Cela est proche de la valeur que l'on s'attend à trouver si deux molécules sont composées au hasard et n'ont aucune relation de parenté. Cette remarque est importante pour bien faire saisir ultérieurement, au niveau moléculaire, les notions d'allèle et de gènes homologues (similitude très nettement supérieure à 25 %).

Variabilité de l'information génétique

Il s'agit ici de relier la variabilité phénotypique et la variabilité de l'ADN, de façon à conforter l'idée que l'information génétique réside dans la séquence de nucléotides du gène.

Couleur des colonies de la Levure Saccharomyces cerevisiae *et variabilité du gène ade2*

Informations scientifiques

Des informations précises concernant le phénotype couleur des colonies de Levures sont disponibles sur le site de Didier Pol : http://www.didier-pol.net/4MUT-LEV.html.

La couleur des colonies

Chez cette Levure, on connaît des souches qui diffèrent par la couleur des colonies qu'elles forment sur milieu solide, colonies blanc-crème pour la souche sauvage, colonies rose-rouge pour une souche mutante.

Cette différence phénotypique a pu au préalable être mise en évidence expérimentalement : c'est un excellent support pour introduire la notion d'hérédité cellulaire.

Cette différence macroscopique entre les deux souches de Levures est due à une différence biochimique en rapport avec la capacité à synthétiser ou non de l'adénine à partir de précurseurs présents dans le milieu. La chaîne de biosynthèse de l'adénine est très complexe et comprend de très nombreuses étapes. Chez la souche mutante à colonies rouges, elle est interrompue à une étape où le produit intermédiaire formé (AIR) est de teinte rose (en milieu aérobie). C'est son accumulation dans une cellule de Levure qui confère à celle-ci une couleur légèrement rose (la couleur rouge de la colonie est due à un effet de masse).

NB : L'adénine (constituant fondamental de l'ATP, des ARN et de l'ADN) est indispensable pour la croissance des Levures, et donc la formation de colonies. Si le milieu est très riche en adénine, la chaîne de biosynthèse est réprimée. Dans ce cas, les Levures mutantes ont le même phénotype que les Levures sauvages, donc blanc-crème. Si le milieu contient de l'adénine en quantité modérée (c'est le cas dans la plupart des milieux standard de culture des Levures), la formation de colonies de Levures mutantes est possible sans que la chaîne de biosynthèse de l'adénine soit réprimée. En conséquence, les Levures mutantes croissent, se multiplient tout en accumulant le produit intermédiaire, d'où la couleur rouge.

Cette observation pourrait éventuellement être utilisée en classe de première comme support pour dégager l'idée que le phénotype dépend de l'interaction du génotype et de l'environnement.

Les allèles du gène ADE2

Le gène ADE2 code pour une enzyme qui transforme le produit intermédiaire coloré (AIR) en un composé CAIR non coloré. Le blocage de cette étape est dû à une mutation dans le gène ADE2. Les séquences de l'allèle sauvage et d'allèles mutés du gène ADE2 ont été établies. La banque de séquences d'*Anagène* fournit pour ADE2 uniquement deux séquences : celle de l'allèle sauvage (ADE2Allele1.adn) et celle d'un allèle muté (ADE2Allele2.adn). La séquence codante de l'allèle muté diffère de l'allèle sauvage au nucléotide 103 (G103T) ce qui transforme le triplet GAA de l'allèle sauvage en TAA dans l'allèle muté (GAA35TAA).

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Variabilité de l'information génétique** permet d'atteindre **Phénotypes couleur des Levures (Gène ADE2)** qui charge le fichier ade2.edi affichant deux séquences : Ade2Allele1 et Ade2Allele2.

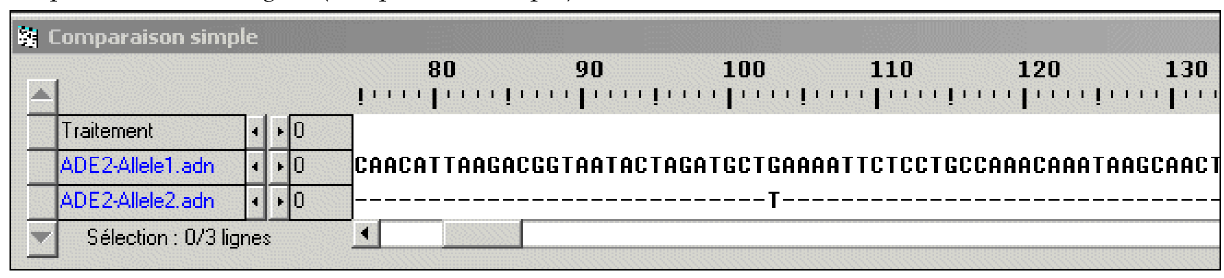
Documents fournis

Dans la banque de documents, le développement de l'arborescence **Variabilité de l'information génétique** permet d'atteindre **Phénotypes couleur des Levures (Gène ADE2)** pour charger les fichiers colonie-rouge.jpg et colonie-blanc-creme.jpg.

Pistes d'exploitation pédagogique des données et documents fournis

Les documents fournis (colonie-rouge.jpg et colonie-blanc-creme.jpg) permettent de mettre en évidence les différences phénotypiques au niveau de la colonie et, par là, au niveau cellulaire. Il s'agit alors de mettre ces différences en relation avec les deux allèles déterminant ces phénotypes.

La comparaison avec *Anagène* (comparaison simple) donne le résultat suivant :



Rappel : attention, avec *Anagène* un seul brin est pris en compte (le brin non transcrit).

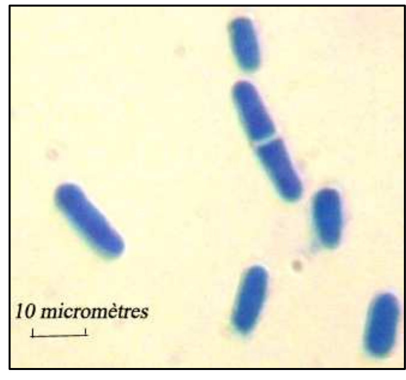
Dans ce cas, on constate qu'un seul changement d'un nucléotide modifie l'information codée par le gène. La composition des deux allèles en nucléotides étant quasiment la même, cela conforte l'idée que l'information génétique réside dans la séquence des nucléotides du gène.

Taille des Levures Schizosaccharomyces pombe *et variabilité du gène cdc2*

Informations scientifiques

Des informations concernant la Levure *Schizosaccharomyces pombe*, ses caractéristiques, ses conditions de culture, sont disponibles sur le site de Didier Pol : http://www.didier-pol.net/4pombe.htm .

Des documents et données complémentaires sur le gène CDC2 se trouvent dans le dossier CDC2 de terminale.



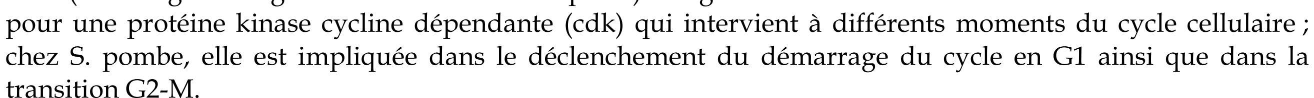
La Levure Schizosaccharomyces pombe

La Levure S. *pombe* se caractérise par ses cellules allongées et cylindriques. Son génome a été entièrement séquencé, et il comprend 4 824 gènes.

Cette Levure, aussi appelée Levure fissipare, ne se multiplie pas par bourgeonnement, mais par fission transversale.

Le gène cdc2

Plusieurs gènes contrôlent le cycle cellulaire, notamment le gène cdc2 (homologue des gènes cdk1 des autres espèces). Ce gène code pour une protéine kinase cycline dépendante (cdk) qui intervient à différents moments du cycle cellulaire ; chez S. pombe, elle est impliquée dans le déclenchement du démarrage du cycle en G1 ainsi que dans la transition G2-M.



La fonction générale des protéines cdk consiste à utiliser l'ATP pour modifier par phosphorylation l'activité de protéines cibles indispensables au bon déroulement du cycle. Ces kinases sont elles-mêmes régulées de façon comparable par phosphorylation et déphosphorylation assurées par d'autres kinases et phosphatases selon des voies régulatrices complexes.

Des mutations qui affectent le gène cdc2 et qui ont pour effet des anomalies de la division cellulaire

Les mutations du gène cdc2 peuvent perturber le déroulement du cycle cellulaire, et donc se traduire phénotypiquement, par exemple par une différence de taille des cellules, ce qui permet d'identifier les mutants :

- mutation cdc2-33 : elle a pour effet de bloquer un signal nécessaire à la transition G2-M, et perturbe donc la division cellulaire sans affecter la croissance de la cellule. Elle se traduit par des cellules de taille beaucoup plus importante que la normale. D'autre part, cette mutation est conditionnelle et ne s'exprime qu'à une température supérieure à 35 °C. On parle alors de mutation thermosensible (des mutations létales peuvent être ainsi conservées tout en laissant la possibilité d'étudier leurs effets phénotypiques) ;

- mutation cdc2-3w (ou wee) : elle a pour effet la production précoce d'un signal d'entrée en mitose. Il y a donc division prématurée et les cellules sont beaucoup plus petites que la normale (*wee* signifie « petit » en écossais).

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Variabilité de l'information génétique** permet d'atteindre **Phénotypes taille des Levures (Gène CDC2)** qui charge le fichier alleles-cdc2-seconde.edi affichant trois séquences : cdc2sac-S.adn, cdc2sac-TS.adn et cdc2sac-DP.adn.

Documents fournis

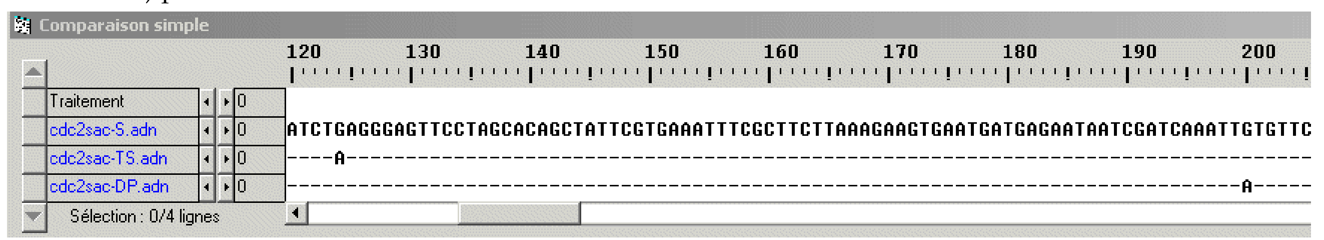
Dans la banque de documents, le développement de l'arborescence **Variabilité de l'information génétique** permet d'atteindre **Phénotypes taille des Levures (Gène CDC2)** pour charger les fichiers SP-division-souche-sauvage.bmp et SP-tailles-3souches.bmp.

Pistes d'exploitation pédagogique des données et documents fournis

Les documents fournis permettent de mettre en évidence les différences phénotypiques en ce qui concerne la taille des trois souches A, B et C : la souche A est une souche sauvage, la souche B est thermosensible et ne se divise pas à 35°, la souche C se divise prématurément à une taille de 7 microns. Le document SP-division-souche-sauvage.bmp est un document de référence relatif à la souche A, qui indique que la taille des Levures est liée à la division cellulaire et varie entre 7 et 14 microns.

Il s'agit alors de mettre en relation les différences phénotypiques de taille entre ces trois souches avec les différences entre les allèles du gène CDC2 qui contrôle l'entrée en division cellulaire.

La comparaison avec *Anagène* (comparaison simple) des séquences des trois allèles fournis (alleles-cdc2-seconde.edi) permet de mettre en évidence les différences suivantes :



Comme pour le gène ADE2, on constate qu'un seul changement d'un nucléotide modifie l'information codée par le gène ; de plus, le site où a lieu le changement en conditionne l'effet puisque les deux allèles entraînent des phénotypes opposés.

La composition des trois allèles en nucléotides étant quasiment la même, cela conforte l'idée que l'information génétique réside dans la séquence des nucléotides du gène.

Phénotypes drépanocytaire ou non drépanocytaire et variabilité du gène de l'hémoglobine

Informations scientifiques

(Des données concernant la drépanocytose sont fournies dans le dossier Relations génotype/phénotype pour la classe de première page 51.)

Les allèles du gène HBB

Au niveau d'une classe de seconde où l'on n'envisage pas le phénotype au niveau protéique, on peut, en première approximation, considérer le gène HBB qui code pour la chaîne bêta de globine humaine, comme un gène qui dirige la synthèse de l'hémoglobine. La banque fournie comprend les séquences de trois allèles dont deux codent pour une chaîne bêta fonctionnelle et le troisième pour une chaîne bêta drépanocytaire.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Variabilité de l'information génétique** permet d'atteindre **Phénotypes drépanocytaires chez l'Homme (Gène HBB)** qui charge le fichier alleles-HB-Beta.edi affichant trois séquences : Hbb nucléique, Hbb nucléique bis et HbS nucléique.

Documents fournis

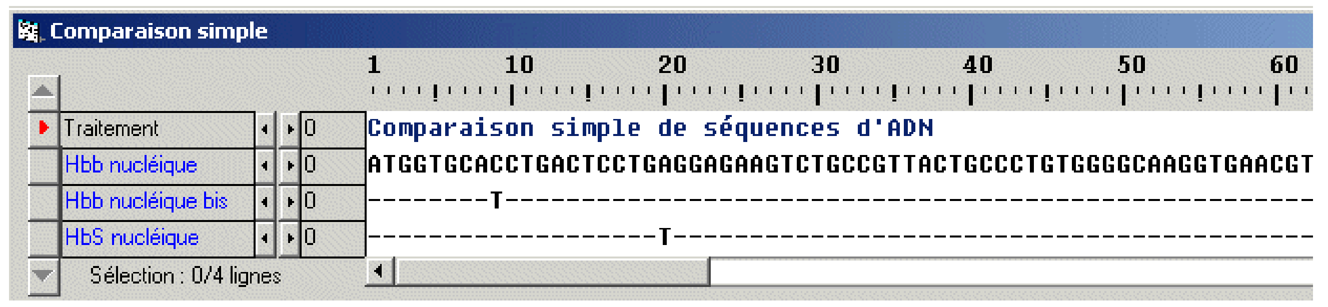
Dans la banque de documents, le développement de l'arborescence **Variabilité de l'information génétique** permet d'atteindre **Phénotypes drépanocytaires chez l'Homme (Gène HBB)** pour charger les fichiers hematies.bmp et phenodrep.bmp.

Pistes d'exploitation pédagogique des données et documents fournis

Les documents fournis permettent de comparer les deux phénotypes au niveau clinique et cellulaire, et de relier ces deux niveaux de formulation du phénotype.

L'étude de cet exemple permet de voir si chez un organisme pluricellulaire, une différence clinique macroscopique peut être reliée à une différence dans la séquence de nucléotides du gène qui intervient dans la réalisation du caractère. En reliant le phénotype macroscopique au phénotype cellulaire, il illustre le fait qu'un gène n'intervient qu'au niveau des cellules. Il n'est pas nécessaire d'entrer dans la structure moléculaire des chaînes des globines. Pour un élève de seconde, ce que régit le gène, c'est la production d'une hémoglobine qui reste soluble dans les hématies ou d'une hémoglobine qui forme des fibres.

La comparaison avec *Anagène* (comparaison simple) des séquences de nucléotides des allèles normal et drépanocytaire du gène de la bêta globine donne le résultat suivant :



Cet exemple conforte l'idée que, chez tous les organismes, une différence phénotypique est toujours associée à une différence des nucléotides de la séquence du gène. En plus, il révèle que l'inverse n'est pas vrai, c'est-à-dire qu'une différence dans la séquence n'entraîne pas obligatoirement une différence phénotypique. Cette remarque sera importante quand on abordera l'idée que des gènes présents chez des espèces différentes, peuvent être considérés comme étant le même gène (gènes homologues) malgré des différences relativement nombreuses dans leurs séquences.

Origine de la variabilité de l'information génétique

Il s'agit d'expliquer la diversité allélique mise en évidence précédemment et d'arriver à l'idée qu'un allèle provient d'un autre allèle à la suite d'un événement appelé mutation.

L'approche expérimentale à partir d'organismes unicellulaires comme les Levures (apparition de colonies blanc-crème à partir d'une souche de Levures rouges soumise à un rayonnement UV. Voir site de Didier Pol) est sans doute le meilleur moyen d'introduire la notion de mutation. Elle présente l'avantage de faire saisir à l'élève que la mutation est fondamentalement un phénomène cellulaire qui modifie la séquence des nucléotides d'un gène.

Ensuite, chez des organismes pluricellulaires, il s'agit de montrer que la mutation, toujours un phénomème cellulaire, peut affecter soit des cellules somatiques, soit des cellules de la lignée cellulaire conduisant à la production des gamètes et que seules les mutations germinales peuvent se transmettre à la génération suivante.

Pour cela, l'exemple des mutations affectant le gène P53, gène dit suppresseur de tumeurs, est un support adapté.



Étude de cancers impliquant des mutations somatiques

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Origine de la variabilité de l'information génétique** permet d'atteindre **Références cancer œsophage** pour charger le fichier ref-cancer-Œsophage.edi qui affiche des séquences strictement codantes des allèles du gène de la p53 présents dans une cellule normale (cn) et dans une cellule cancéreuse (cc) chez un individu atteint d'un cancer de l'œsophage.

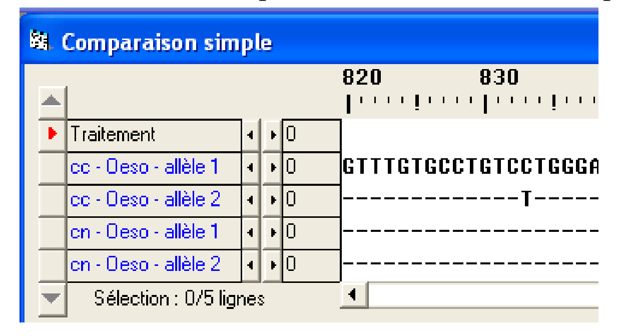
Documents fournis

Dans la banque de documents, le développement de l'arborescence **Origine de la variabilité de l'information génétique** permet d'atteindre **Mutations somatiques et mutations germinales** pour charger les fichiers infocance.bmp (texte de présentation des phénotypes cancéreux au niveau cellulaire et clinique) et mutations.bmp (texte précisant la différence entre mutation germinale et mutation somatique).

Pistes d'exploitation pédagogique des données fournies

Le document infocancer.bmp fournit des renseignements sur le comportement des cellules cancéreuses et sur l'implication du gène p53 dans ce comportement. À partir de là, il s'agit d'exploiter les données fournies par les séquences pour construire la notion de mutation somatique.

Le document ci-dessous présente le résultat de la comparaison effectuée dans le cas du cancer de l'œsophage :



les cellules cancéreuses (ccOeso) possèdent deux allèles différents du gène p53 : un allèle normal (allèle 1) et un allèle muté (allèle 2) ; les deux allèles diffèrent par une substitution d'un nucléotide en position 833 ;

les cellules normales du même individu possèdent deux allèles normaux du gène p53.

Il y a donc une différence de génotype entre les cellules cancéreuses et les cellules normales. Cette différence explique la différence de phénotype cellulaire : cellule cancéreuse qui se divise de façon anarchique et cellule normale dont la division est contrôlée et limitée.

L'élève devrait conclure que c'est dans la cellule initiale à l'origine de la tumeur qu'a eu lieu la mutation transformant un allèle P53 normal en un allèle muté. Et à l'aide du document mutations.bmp conclure qu'il s'agit ici d'une mutation somatique.

La même démarche et les mêmes conclusions peuvent être obtenues à partir de l'exemple du cancer du foie. On observe alors une mutation par substitution en position 747 : G remplacé par T, uniquement dans les cellules cancéreuses.

Étude d'un cancer impliquant une mutation germinale (syndrome de Li-Fraumeni, qui se traduit par des cancers multiples et précoces)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Origine de la variabilité de l'information génétique/Mutations somatiques et mutations germinales** permet d'atteindre :

- **Référence cancer LiFraumeni** qui charge le fichier ref-cancer-LiFrau.edi affichant des séquences strictement codantes des allèles du gène p53 présents dans une cellule normale (cn) et dans une cellule cancéreuse (cc) chez un individu atteint du syndrome de Li-Fraumeni ;

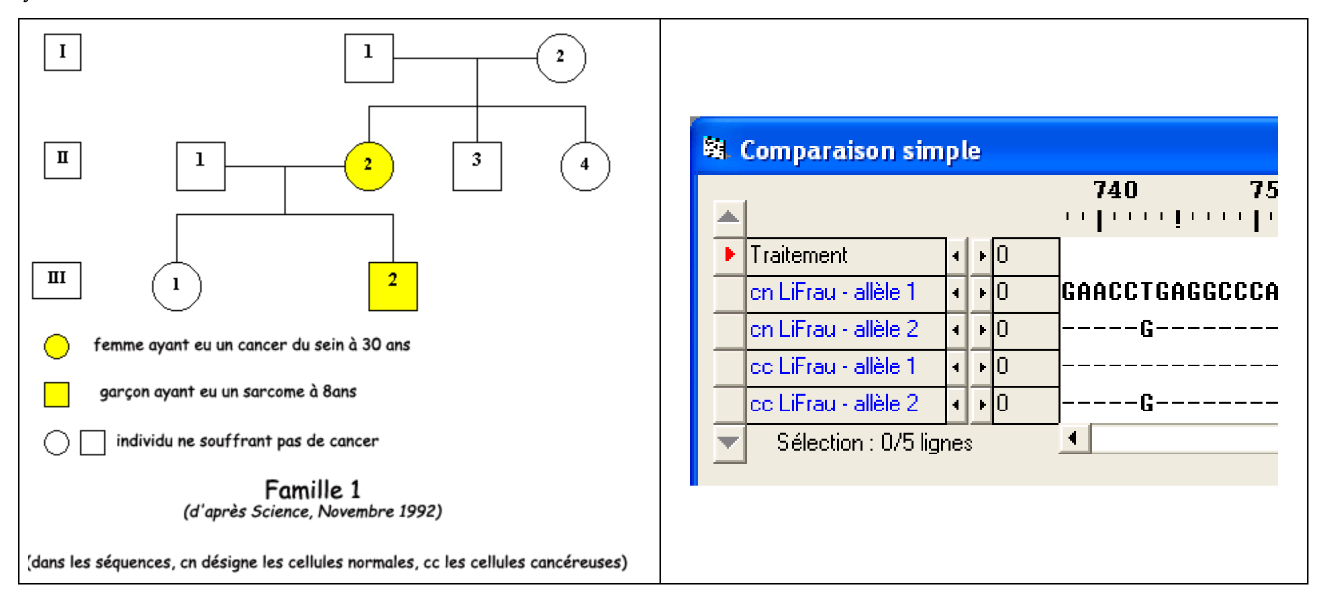
- **Famille1 P53** qui charge le fichier Famille1-P53.edi affichant les séquences strictement codantes des allèles du gène p53 présents dans les cellules de chacun des membres de la famille 1. Si l'individu est atteint d'un cancer, les séquences sont fournies pour les cellules normales et pour les cellules cancéreuses. Le nom de séquence I1 – cn – allèle 1 signifie qu'il s'agit de l'allèle 1 du gène p53 présent dans une cellule normale de l'individu I1.

Documents fournis

Dans la banque de documents, le chemin **Origine de la variabilité de l'information génétique/Mutations somatiques et mutations germinales** permet d'atteindre **Famille1** qui charge le fichier Famille1-P53.bmp affichant l'arbre généalogique de la famille 1 dans laquelle deux individus sont atteints du syndrome de Li-Fraumeni. Il s'agit dans cette famille d'une mutation germinale apparue de novo chez l'individu II2.

Pistes d'exploitation pédagogique des données fournies

Pour aborder cette étude, il faut partir de l'arbre généalogique qui montre deux personnes, II2 et III2, étroitement apparentées qui ont un cancer à un âge précoce. En outre, l'arbre suggère une transmission héréditaire. Ces caractéristiques « observation d'un sarcome chez un sujet atteint de moins de 45 ans, apparenté au premier degré à une personne ayant eu un cancer de n'importe quel type avant 45 ans, ou au deuxième degré à une personne ayant eu un cancer ou un sarcome à moins de 45 ans » définissent le syndrome de Li-Fraumeni.



La comparaison des allèles du gène p53 présents dans les cellules cancéreuses et dans les cellules normales des personnes atteintes de cancer (II2 et III2) (« LiFrau.edi ») avec l'allèle P53 de référence, montre que toutes les cellules possèdent deux allèles différents : un allèle normal et un allèle muté. L'élève doit en conclure que la cellule-œuf à l'origine de l'organisme possédait également un allèle normal et un allèle muté.

		740 750 7
Traitement	0	
I1 - cn - allèle 1	0	CGGCATGAACCGGAGGCCCATCCTCACCA
I1- cn - allèle 2	0	-----------------------------
I2 - cn - allèle 1	0	-----------------------------
I2 - cn - allèle 2	0	-----------------------------
II1 - cn - allèle 1	0	-----------------------------
II1 - cn - allèle 2	0	-----------------------------
II2 - cc - allèle 1	0	-----------------------------
II2 - cc - allèle 2	0	-----------T-----------------
II2 - cn - allèle 1	0	-----------------------------
II2 - cn - allèle 2	0	-----------T-----------------
II3 - cn allèle 1	0	-----------------------------
II3 - cn - allèle 2	0	-----------------------------
II4 - cn - allèle 1	0	-----------------------------
II4 - cn - allèle 2	0	-----------------------------
II5 - cn - allèle 1	0	-----------------------------
II5 - cn - allèle 2	0	-----------------------------
III1 - cn - allèle 1	0	-----------------------------
III1 - cn - allèle 2	0	-----------T-----------------
III2 - cc - allèle 1	0	-----------T-----------------
III2 - cc - allèle 2	0	-----------------------------
III2 - cn - allèle 1	0	-----------T-----------------
III2 - cn - allèle 2	0	-----------------------------

Sélection : 0/23 lignes

Remarque : le fait que des cellules normales possèdent un allèle muté de P53 peut être utilisé pour faire saisir que le phénotype cellule cancéreuse ne dépend pas du seul gène P53. La possession d'un allèle muté de P53 prédispose la cellule à devenir cancéreuse.

Il faut s'interroger sur l'origine de l'allèle muté présent dans toutes les cellules de l'organisme des individus II2 et III2. La première conclusion est que III2 l'a hérité de II2. Pour saisir l'origine de l'allèle muté chez l'individu II2, il faut étudier les génotypes de I1 et I2. La comparaison avec l'allèle de référence de P53 indique que ces deux personnes possèdent uniquement l'allèle de référence (normal). À partir de là, et en s'aidant éventuellement du document (mutations.bmp), l'élève doit conclure que la présence de l'allèle muté chez II2 résulte d'une mutation germinale survenue chez un des deux parents. En toute rigueur, on ne peut exclure totalement une mutation survenue au stade œuf.

Si l'on veut pousser l'analyse plus loin, on constate que la personne III1 possède un allèle muté qu'elle a hérité de sa mère, mais n'a pas de cancer.

Ainsi, il faut prendre soin de repréciser que le cancer est une maladie plurifactorielle et que les descendants de personnes atteintes d'un cancer dû à une mutation germinale ne seront pas forcément atteints aux aussi par un cancer ; simplement, leur risque est plus important, d'où l'intérêt d'exercer une vigilance accrue.

Détermination du type de mutation (germinale ou somatique) impliquée dans une famille touchée par le cancer : mutations et environnement

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Origine de la variabilité de l'information génétique/Mutations somatiques et mutations germinales** permet d'atteindre :

- **Références cancer foie** pour charger le fichier ref-cancer-Foie.edi qui affiche des séquences strictement codantes des allèles du gène p53 présents dans une cellule normale (cn) et dans une cellule cancéreuse (cc) chez un individu atteint d'un cancer du foie ;

- **Famille2 P53** pour charger le fichier Famille2-P53.edi qui affiche les séquences strictement codantes des allèles du gène p53 présents dans les cellules de chacun des membres de la famille 2. Si l'individu est atteint d'un cancer, les séquences sont fournies pour les cellules normales et pour les cellules cancéreuses.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Origine de la variabilité de l'information génétique** permet d'atteindre **Mutations somatiques et mutations germinales** pour charger les fichiers :

- famille2.bmp : dans l'arbre généalogique, trois individus sont atteints d'un cancer du foie. Il s'agit ici d'une mutation somatique. Sa fréquence élevée s'explique par l'influence d'un facteur de l'environnement, l'aflatoxine ;

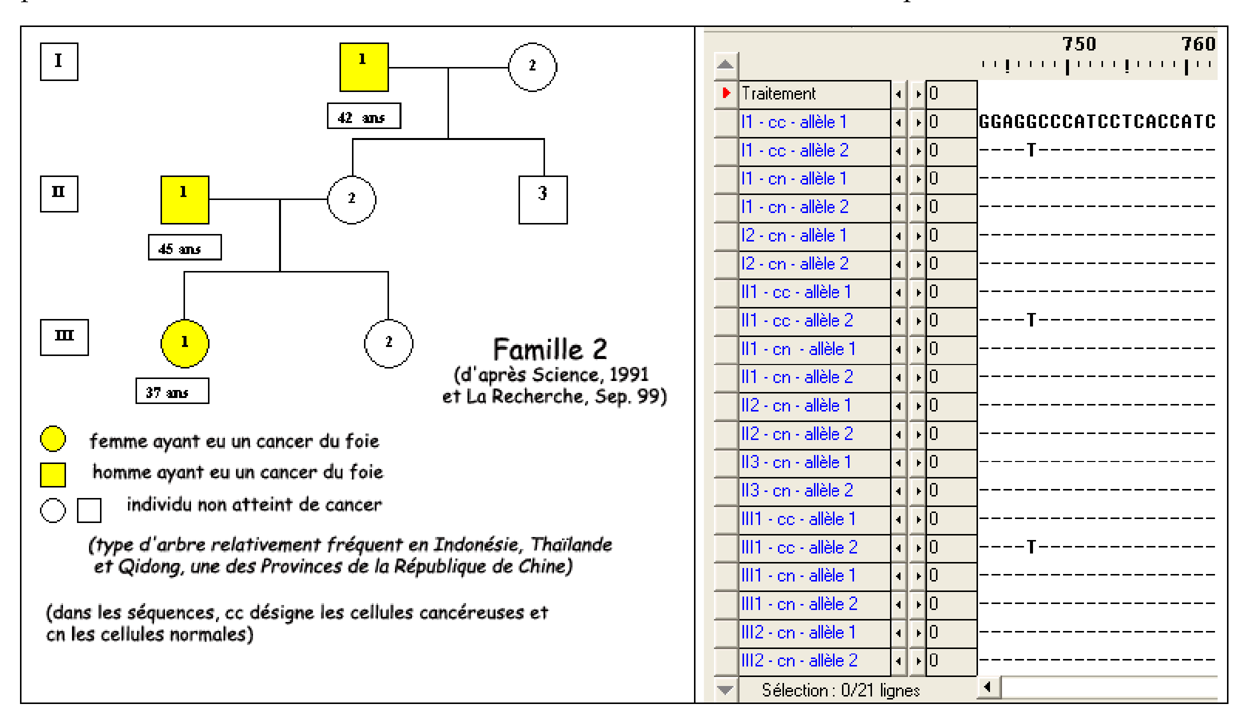
- aflatoxine.bmp : texte mettant en relation un facteur de l'environnement, l'aflatoxine, avec le phénotype « cancer du foie ».

Pistes d'exploitation pédagogique des données fournies

La comparaison des allèles présents dans la cellule normale et de ceux présents dans la cellule cancéreuse permet de déterminer l'origine du cancer. On arrive ainsi aux résultats et conclusions suivantes :

- **Dans cette famille 2**, un individu est touché à chaque génération (I1, II1, III1). Seules les cellules cancéreuses de ces personnes possèdent un allèle muté et un allèle normal ; leurs cellules normales possèdent deux allèles normaux. On a donc affaire à des mutations somatiques ;

.



- on peut alors s'interroger sur la fréquence élevée de ce cancer du foie dans cette famille, alors qu'il n'est pas héréditaire. Une information supplémentaire est apportée par l'arbre : ce type d'arbre est fréquent dans certaines populations. On peut alors supposer l'influence d'un facteur de l'environnement. Un document (aflatoxine.bmp) apporte alors des informations complémentaires : c'est une toxine, l'aflatoxine, libérée par un champignon souvent présent dans la nourriture de ces populations, qui est responsable de ces cancers.

On peut ainsi illustrer l'idée qu'un facteur de l'environnement peut influencer un phénotype.

Unité du vivant

Division cellulaire chez les êtres vivants et gène cdc2

Les caractéristiques communes à tous les êtres vivants précédemment étudiés (généralité de la structure cellulaire, information génétique codée dans des molécules d'ADN de même structure), permettent d'évoquer leur origine commune. La recherche de gènes partagés par les organismes peut renforcer cette idée. Dans cette optique, le gène CDC2 envisagé précédemment est intéressant car des gènes homologues sont présents chez tous les Eucaryotes.

Séquences et documents

Fichiers des séquences

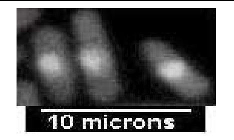
Dans la banque des thèmes d'étude, le développement de l'arborescence **Unité du vivant** permet d'atteindre **Division cellulaire chez plusieurs Eucaryotes (Gène CDC2)** pour charger le fichier evol-cdc2-adn.edi affichant des allèles du gène CDC2 chez plusieurs êtres vivants.

Documents fournis

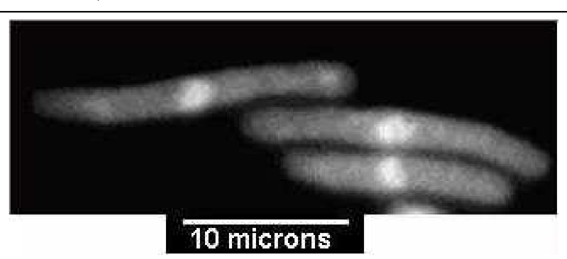
Dans la banque de documents, le développement de l'arborescence **Unité du vivant** permet d'atteindre **Division cellulaire chez plusieurs Eucaryotes (Gène CDC2)** pour charger deux fichiers : transgeneseSP.jpg (expérience de transgenèse) et Division-SP.bmp (Levure SP en division).

Pistes d'exploitation pédagogique

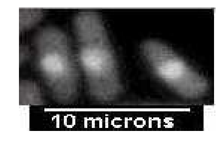
L'étude du document **Expérience de transgenèse** (éventuellement associé au document **Levure SP** en division) montre que l'insertion du CDC2 humain dans une cellule de Levure fissipare dont le gène CDC2 a été lésé, rétablit la capacité de la cellule à se diviser (figure en bas à droite du document). Ceci indique une similitude fonctionnelle entre les gènes CDC2 humain et de Levure, traduisant l'unité du monde vivant.



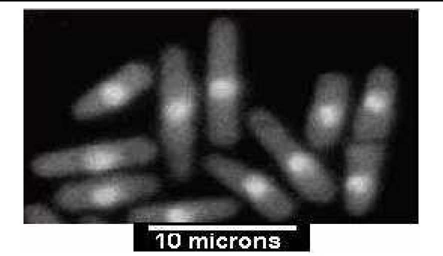
Levures de souche sauvage dont le gène CDC2 a été lésé, cultivées à 25 °C



Levures douze heures plus tard



Levures de souche sauvage dont le gène CDC2 a été lésé, et dans lesquelles on a transféré le gène CDC2 humain. Culture à 25 °C



Levures douze heures plus tard

Expérience de transgenèse (schématisée et adaptée d'après M.G. Lee et P. Nurse, Nature*, 1987, 327, 31-35)*

Cette perception de l'unité du vivant peut être renforcée par la comparaison des gènes CDC2 de différentes espèces. Pour obtenir le nombre d'identités et de similitudes entre des séquences comparées, il faut demander l'alignement avec discontinuités, cliquer sur l'icône avec un « I » qui donne l'information.

	Homme	Rat	Xénope	Drosophile	Arabidopsis	Levure
Homme	0	84,6	73,9	62,4	63,7	60,7
Rat		0	73,6	62,9	62,2	57,9
Xénope			0	63,4	60,9	59,5
Drosophile				0	59,6	57,3
Arabidopsis					0	60,5
Levure						0

*Identités en % obtenues avec l'alignement multiple d'*Anagène

La similitude entre les gènes CDC2 est supérieure à 60 %. Associée à l'expérience de transgenèse, elle conduit à l'idée que tous ces organismes possèdent en commun ce gène, héritage d'un gène CDC2 possédé par l'ancêtre commun à toutes ces espèces.

Remarques : Pour un élève, des similitudes de l'ordre de 60 % peuvent paraître non significatives. Il faut donc rappeler ce qui a été vu au moment de l'étude de l'ADN, à savoir que si deux séquences n'ont aucun lien entre elles, la similitude est de l'ordre de 25 %.

En outre, il peut paraître bizarre pour les élèves que des différences nombreuses entre les gènes n'entraînent pas de conséquences fonctionnelles, alors qu'une différence réduite entre deux allèles avait dans les exemples précédents des conséquences phénotypiques importantes.

Il y a donc nécessité de rappeler que si une différence fonctionnelle est toujours liée à une différence dans la séquence, la réciproque n'est pas vraie (exemple vu à propos d'un allèle de HBB).

Plans d'organisation des animaux et gènes homéotiques

Informations scientifiques

Ces informations concernent les thèmes susceptibles d'être traités en seconde et terminale. En revanche, les suggestions pédagogiques qui figurent dans ce chapitre sont propres à la classe de seconde.

À propos des gènes homéotiques

Les mutants homéotiques

Les phénotypes homéotiques ont été décrits pour la première fois en 1894 par William Bateson, qui utilisa le terme « homeosis » pour désigner le phénomène de transformation d'un organe en un autre. C'est l'étude de ces mutants présentant des anomalies de disposition de certains organes (mutants homéotiques) qui a été à l'origine de la découverte des gènes homéotiques. Un gène homéotique est alors défini comme un gène dont la mutation produit une homéose, c'est-à-dire l'apparition d'un organe bien formé mais à un mauvais emplacement dans le corps. Les documents ci-dessous présentent quelques exemples de phénotypes mutants homéotiques chez la Drosophile :

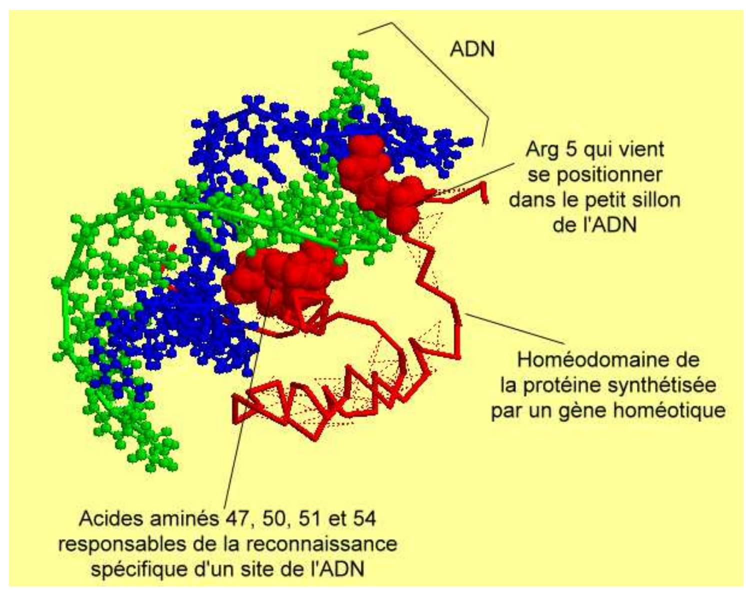
Mutant Drosophile antennapedia *Présence de pattes à la place des antennes*	
Mutant Drosophile bithorax (bx, découvert en 1941) *Transformation du 3e segment thoracique en 2e segment thoracique – présence de deux paires d'ailes*	
Mutant Drosophile bithoraxoïde (bxd, découvert en 1927) *Transformation du premier segment abdominal en 3e segment thoracique – présence de quatre paires de pattes*	
Mutant Drosophile ultrabithorax (ubx) *Transformation du 3e segment thoracique et du 1er segment abdominal en 2e segment thoracique (létal à l'état larvaire)*	Larve normale — Larve mutante Ubx

Les gènes homéotiques

Certains gènes vont intervenir au cours du développement embryonnaire, notamment pour diriger l'expression d'autres gènes et permettre ainsi la mise en place des différents organes ; on les qualifie de gènes « maîtres » ou gènes « architectes ». Il en est ainsi chez les Vertébrés des gènes Pax, qui déterminent la disposition des organes selon l'axe dorso-ventral, et des gènes homéotiques qui déterminent la disposition des organes le long de l'axe antéro-postérieur.

Les gènes homéotiques sont des gènes régulateurs (activation ou inhibition) de l'expression d'autres gènes.

Ce sont des gènes à homéoboîte (ou homeobox), c'est-à-dire qu'ils codent pour une protéine dont un domaine est très conservé au cours de l'évolution. Ce domaine protéique, appelé homéodomaine, a été découvert en 1983 ; il comporte soixante acides aminés et sa structure secondaire comporte trois hélices formant le motif HLH (Helix Loop Helix). Ce domaine se fixe à l'ADN, ce qui contribue à ouvrir localement l'ADN pour permettre la transcription.



Complexe Homéodomaine/ADN (cette figure a été obtenue avec le logiciel RasTop)

Les gènes homéotiques sont organisés en complexes homéotiques chez les animaux.

Les complexes homéotiques

La disposition et l'organisation des gènes homéotiques sont bien connues chez deux organismes : la Drosophile et la Souris.

Chez la Drosophile, les gènes homéotiques forment le complexe Hom-C porté par le chromosome 3 ; ce complexe est composé de deux groupes de gènes : le groupe Antennapedia et le groupe Bithorax, chacun de ces groupes comprenant plusieurs gènes :

- Antennapedia : gènes lab (labial), proboscipedia (pb), Deformed (dfd), Sex comb reduced (Scr), Antennapedia (antp) ;
- Bithorax : gènes ultrabithorax (ubx), abdominal A (AbdA) et abdominal B (AbdB).

Chez la Souris, les gènes homéotiques forment quatre complexes Hox répartis sur quatre chromosomes :

- complexe des gènes Hox A sur le chromosome 6 ;
- complexe des gènes Hox B sur le chromosome 11 ;
- complexe des gènes Hox C sur le chromosome 15 ;
- complexe des gènes Hox D sur le chromosome 2.

Il y a colinéarité entre la disposition des gènes homéotiques sur les chromosomes et les régions du corps dans lesquelles ils s'expriment : en parcourant l'ADN de 3′ vers 5′, on trouve des gènes dont les lieux d'action s'échelonnent de l'avant vers l'arrière de l'animal.

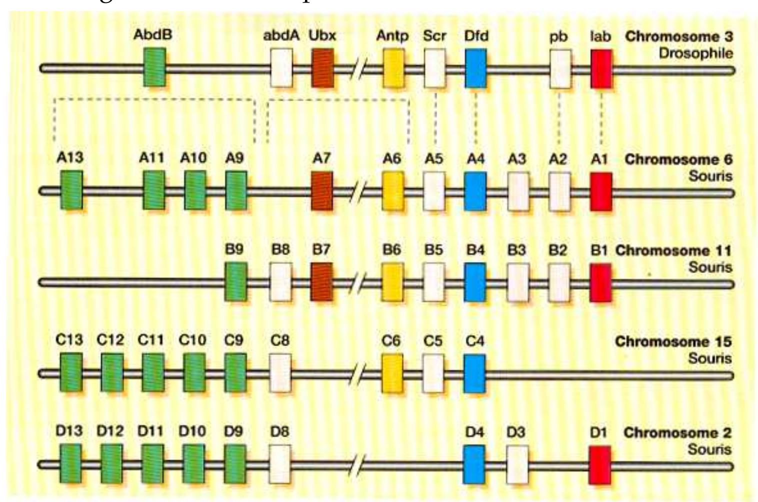
Lors de l'embryogenèse, la combinaison des produits de plusieurs gènes Hox donne une identité relative aux cellules embryonnaires le long de l'axe antéro-postérieur. Chaque cellule embryonnaire est affectée d'une valeur positionnelle, qui résulte d'une combinatoire de plusieurs gènes Hox.

Les gènes homéotiques des Arthropodes et des Vertébrés forment une famille multigénique

Au sein d'un même organisme, par exemple la Souris, tous les gènes homéotiques présentent une similitude importante de séquence au niveau de l'homéoboîte : ce sont donc des gènes homologues, cette homologie témoignant de leur origine commune. De plus, l'ordre des gènes sur chaque chromosome (donc au sein d'un même complexe) est similaire d'un complexe à l'autre. On peut donc dire que l'on a des gènes paralogues (exemple : HoxA4, HoxB4, HoxC4 et HoxD4 sont paralogues).

Si l'on effectue maintenant des comparaisons entre les gènes homéotiques des Vertébrés et des Arthropodes :

- les gènes des complexes Hox de Souris et Hom-C de Drosophile présentent des similitudes dans leur ordre de disposition sur les chromosomes et, dans les deux cas, ils sont impliqués dans le positionnement et l'identité cellulaire le long de l'axe antéro-postérieur ;



- des expériences de transgenèse interspécifique ont été réalisées avec succès :
  - o la substitution du gène lab muté chez la Drosophile (mutation qui entraîne la formation d'une tête très petite) par le gène Hox B1 de Poulet restaure la morphogenèse normale de la tête de Mouche ;
  - o l'insertion du gène Hox B9 de Souris chez la Drosophile entraîne la transformation des antennes en pattes thoraciques ;
  - o d'une manière générale, un gène homéotique de Vertébré possède son homologue chez les Insectes ; inséré par transgenèse interspécifique dans le génome de la Drosophile, le gène homéotique de la Souris ou de l'Homme induit des transformations homéotiques semblables à celles produites par son homologue.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Unité du vivant** permet d'atteindre **Un gène homéotique (Antennapedia - HOXB6) chez des animaux pluricellulaires** pour charger le fichier B6-ANTP.edi qui affiche des séquences nucléiques des homéoboîtes du gène Hoxb6 d'Homme, de Souris, et du gène Antp de Drosophile.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Unité du vivant** permet d'atteindre **Un gène homéotique (Antennapedia – HOXB6) chez des animaux pluricellulaires** pour charger trois fichiers :

- chromosomeshomeo.jpg : schéma comparatif de la disposition chromosomique des gènes homéotiques chez la Drosophile et chez la Souris, et des territoires corporels d'expression des différents gènes ;

- phenotypeshomeo.jpg : photos représentant trois mutants homéotiques de Drosophile (antennapedia, bithorax et bithoraxoïde) ;

- transgenese.jpg : texte de présentation de trois expériences de transgenèse réussies entre Vertébrés et Drosophile.

Pistes d'exploitation pédagogique en classe de seconde

L'exploitation des données sur les homéoboîtes des gènes homéotiques permet d'aborder la notion d'unité du vivant. On se limite volontairement à l'exploitation des données concernant un seul gène homéotique : Hox B6 chez les Vertébrés, et son homologue Antp chez la Drosophile (ne sont fournies ici que les séquences des homéoboîtes). En relation avec les dissections réalisées et la découverte de la notion de plan d'organisation, l'exploitation des documents fournis permettra de préciser le déterminisme de la mise en place de ces plans d'organisation et de mettre en évidence des ressemblances entre différents groupes (donc une unité du vivant) :

- la description des phénotypes de mutants homéotiques (phenotypeshomeo.jpg) permet de comprendre l'action des gènes homéotiques. Ces mutations entraînent la formation d'un organe parfaitement constitué, mais pas au bon endroit. Les gènes homéotiques sont donc des gènes qui vont intervenir au cours du développement embryonnaire pour permettre la mise en place des différents organes. Ils sont responsables de l'édification du plan d'organisation de l'organisme ;
- la disposition des gènes sur les chromosomes (chromosomeshomeo.jpg) permet de faire plusieurs constats :
  - la disposition des gènes sur les chromosomes correspond à la disposition des régions d'expression de ces gènes dans l'organisme ;
  - il y a une correspondance qui a pu être établie entre les gènes de la Drosophile et les gènes de la Souris (ces correspondances sont indiquées par les couleurs) ;
- les expériences de transgenèse réussies (transgenese.bmp) sont un argument supplémentaire en faveur de l'homologie des gènes homéotiques de Vertébrés et d'Arthropodes, et donc de l'unité du vivant.

La comparaison des séquences nucléiques fournies renforcera l'idée d'unité du vivant et de parentés plus ou moins grandes entre les êtres vivants. Cette comparaison donne les résultats suivants (matrice des identités, valeurs en %) :

	Homme	Souris	Drosophile
Homme	0	92,2	83,3
Souris		0	81,7
Drosophile			0

Toute la difficulté consiste à faire prendre conscience à l'élève que les différences, qui peuvent paraître importantes à première vue, sont finalement très faibles, trop faibles pour envisager l'action du hasard. Ce fort taux d'identités ne peut alors s'interpréter que par une origine commune de ces gènes, donc des animaux qui les possèdent.

Si l'on veut aller un peu plus loin, on peut discuter sur les variations de ce taux d'identités entre les trois espèces : très peu de différences entre les gènes d'Homme et de Souris témoignent d'une parenté plus grande entre eux.

SUGGESTIONS PÉDAGOGIQUES : CLASSE DE PREMIÈRE SÉRIES S, L ET ES

INTRODUCTION : LA BANQUE DE DONNÉES ET LES NOTIONS DU PROGRAMME

Le programme de biologie de la classe de première est centré sur la façon dont le génotype en interaction avec l'environnement, détermine le phénotype d'un organisme. Cela suppose la reconnaissance des divers niveaux de définition du phénotype.

- Dans un premier temps, il s'agit d'établir les mécanismes de base par lesquels des différences entre allèles d'un gène entraînent des différences au niveau du phénotype moléculaire à l'origine des différences des phénotypes aux échelles cellulaire et macroscopique. Ceci implique la connaissance des principes de la synthèse des protéines, du code génétique ainsi que celle du rôle des protéines notamment des enzymes dans la vie cellulaire.

  Plusieurs des exemples fournis dans la banque de données permettent d'une part d'introduire les niveaux de définition du phénotype et le fait qu'un allèle en s'exprimant régit la synthèse d'un polypeptide ayant une séquence d'acides aminés déterminée, ce qui motive l'étude de la synthèse protéique. La connaissance des principes de la synthèse protéique permet ensuite de rechercher et d'expliquer comment une différence allèlique engendre une différence phénotypique.

- Dans un deuxième temps, il s'agit d'aller au-delà d'un déterminisme génétique strict pour envisager la complexité des relations génotype–phénotype–environnement. Ceci fait appel aux notions de dominance/récessivité, au fait que des allèles différents d'un gène peuvent avoir des conséquences phénotypiques variables, que plusieurs allèles interviennent directement ou indirectement dans la réalisation d'un phénotype, et que les facteurs d'environnement, en agissant à différents niveaux du phénotype, peuvent le modifier.

  Les exemples de la banque retenus pour établir les notions de base peuvent servir aussi pour faire saisir la complexité du phénotype macroscopique d'un organisme. Dans le tableau suivant sont spécifiées les différentes notions pouvant être construites à partir de chaque exemple. Dans la mesure du possible, ces exemples ont été construits suivant le même schéma pour faciliter la conception des démarches d'exploitation.

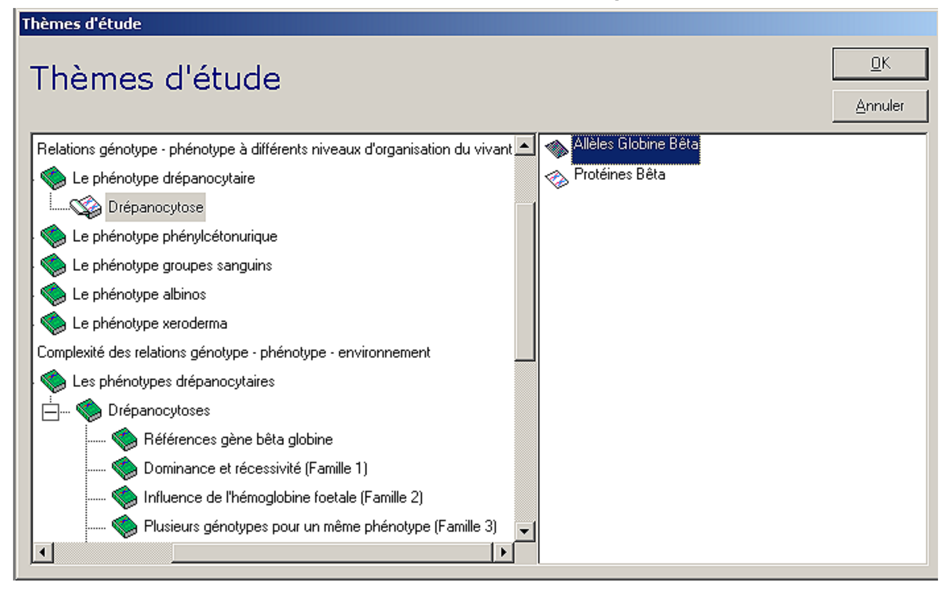
	Rôle de la protéine (phénotype moléculaire)				
Exemples	**Divers niveaux du phénotype**	**Conséquences allèliques**	**Notion de dominance**	**Intervention de deux gènes ou plus**	**Action de l'environnement**
Drépanocytose	X	X	X	Hb bêta et Hb Fœtale	Oxygénation
Phénylcétonurie	X	X	X	Gène de la PAH et gène DHPR	Alimentation
Groupes sanguins ABO	X	X	X	Gènes A, B, O et FUT1	
Albinisme	X	X	X	Gène de la tyrosinase et Gène P	Température
Enphysème pulmonaire	X	X	X	-	Fumée de tabac
Xéroderma	X	X	X		Rayonnement UV
Cancer du sein	X	X	X		Activité physique

Noter que certains de ces exemples seront repris et enrichis en classe de terminale pour envisager d'autres thématiques telles que : innovations génétiques, filiation entre allèles, diagnostic génétique.

Noter également que les deux étapes de cette étude sont présentées successivement pour chaque thème dans cette documentation. Les séquences et les documents s'y rapportant seront à prendre dans deux entrées distinctes de l'arborescence des thèmes :

- relations génotype–phénotype à différents niveaux d'organisation du vivant ;
- Complexité des relations génotype–phénotype–environnement ;

ceci afin de ne pas multiplier les séquences et les documents à charger pour chaque phase du travail.



Le programme prévoit également d'envisager la morphogenèse végétale sous l'angle des interactions génotype–phénotype–environnement. Le déterminisme du phénotype alternatif « nain » et « taille normale » de nombreuses plantes constitue un bon support. La croissance de la plante dépend de la synthèse de gibbérellines, hormones végétales. La chaîne de synthèse des gibbérellines et la séquence d'un gène impliqué dans la genèse du phénotype nain sont connues. En outre, chez les plantes en rosette comme l'Épinard, c'est un facteur de l'environnement, la longueur de la photopériode, qui déclenche l'expression des gènes de cette chaîne de biosynthèse. Cet exemple enrichit ainsi la perception de la complexité des relations génotype–phénotype–environnement puisqu'il sensibilise à l'idée que les facteurs d'environnement peuvent agir directement sur l'expression des gènes.

L'orientation majeure de l'étude de la physiologie nerveuse en première S est la plasticité cérébrale, c'est-à-dire la perception de la modulation de l'organisation cérébrale par les facteurs de l'environnement. Néanmoins, ceux-ci agissent sur un cerveau dont la construction est régie par des milliers de gènes. Ce déterminisme génétique peut être abordé à partir des mutants cérébelleux.

Sur tous ces exemples, des informations complémentaires sont disponibles sur le site *Biotic* de l'INRP : http://www.inrp.fr/Acces/biotic/gpe/accueil.htm

GÉNOTYPE, PHÉNOTYPE, ENVIRONNEMENT

Les phénotypes drépanocytaires

Informations scientifiques

Le phénotype clinique

La maladie se traduit par une anémie et une fatigue permanente, et par la survenue de crises drépanocytaires plus ou moins graves. Ces crises sont dues à des ischémies locales pouvant être très graves ; les crises vaso-occlusives peuvent être particulièrement douloureuses dans les muscles et les risques de complications organiques graves (notamment au niveau du squelette, de la rate, du tube digestif, du cerveau).

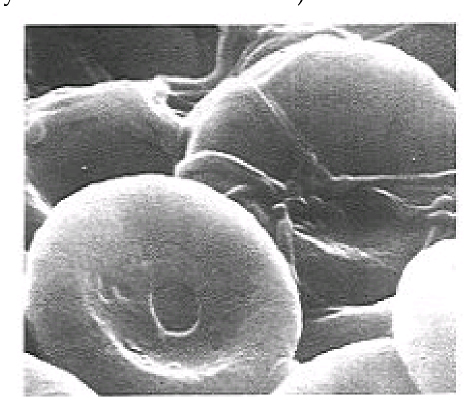
Plusieurs facteurs favorisent la crise drépanocytaire :

- la déshydratation, fréquente chez le drépanocytaire car il est atteint de polyurie ;
- le ralentissement de la circulation sanguine, qui favorise une stase. Il faut donc éviter le port de vêtements trop serrés, une mauvaise position, le froid, la fièvre (formation de protéines inflammatoires), les infections (les globules blancs en excès limitent la circulation des hématies) ;
- toute consommation d'oxygène supplémentaire : les efforts avec essoufflement, les efforts musculaires concentrés sur un muscle ;
- tout ce qui désature l'hémoglobine en oxygène : la vie en altitude (éviter les altitudes supérieures à 2 000 m, et même parfois 1 500 m), les voyages en avion, les écarts de température entre l'air et l'eau (piscine, mer), l'alcool, le tabac.

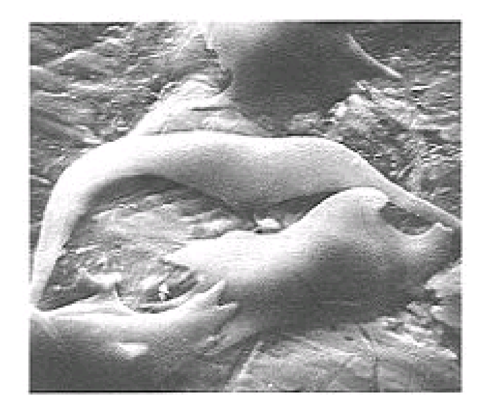
Autrefois, 80 % des individus homozygotes mourraient avant l'âge de la reproduction. Aujourd'hui, grâce au dépistage précoce, à la prévention des infections (vaccination, antibiothérapie systématique), à la prévention de la déshydratation et de toute autre cause pouvant provoquer des troubles chez le malade, la maladie reste sérieuse et invalidante, mais l'espérance de vie s'est considérablement accrue.

Le phénotype cellulaire

Les hématies d'un individu drépanocytaire ont tendance à prendre une forme en faucille (d'où l'autre nom de la drépanocytose : anémie falciforme).



Hématies normales



Hématies drépanocytaires

Cette falciformation est due à la présence de fibres d'hémoglobine S dans l'hématie. Les hématies déformées ralentissent la circulation sanguine dans les capillaires et peuvent même les obstruer.

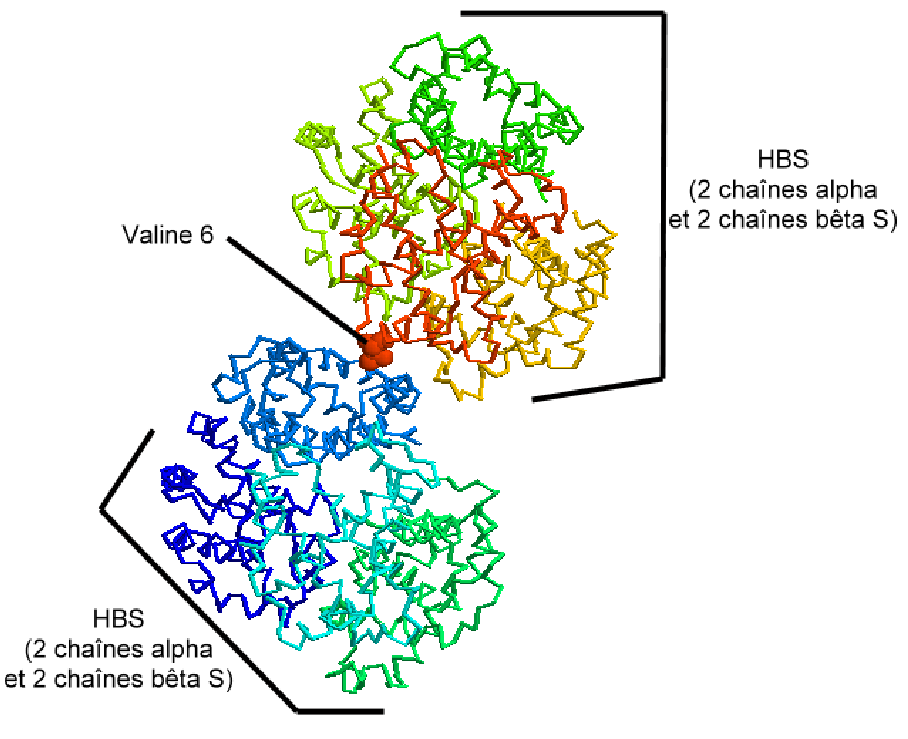
Le phénotype moléculaire

L'hémoglobine S (HbS) est une protéine tétramérique constituée de deux chaînes de globine bêta S et de deux chaînes de globine alpha. La globine bêta S diffère de la globine normale par un seul acide aminé : de la valine remplace un acide glutamique en position 6. Ce changement ne modifie pas la structure spatiale de la globine bêta, et n'affecte pas la poche de l'hème.

La valine est un résidu hydrophobe, qui remplace donc un résidu hydrophile. Les globines étant entourées par un film d'eau, la présence d'un site hydrophobe crée un « point de collage » entre deux molécules

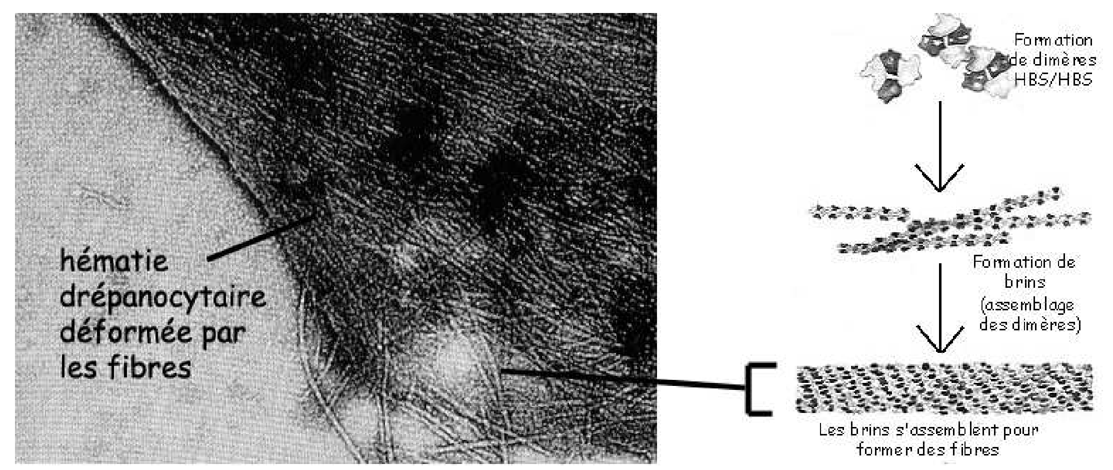
d'hémoglobines voisines ; ce « collage » s'établit entre la leucine 88 et la phénylalanine 85 d'une chaîne ~~alpha~~ bêta et la valine 6 de la chaîne bêtaS. Il se forme alors une structure cristalline en fibres.

La visualisation 3D (obtenue avec le logiciel *RasTop*, fichier « hbshbs.pdb ») ci-dessous met en évidence ce « point de collage » :



Dimère HBS/HBS

Les dimères HbS/HbS s'assemblent pour former des brins ; les brins s'associent en fibres, responsables de la déformation des hématies.



Cette falciformation est due à la formation de fibres d'hémoglobine S (de 1 à 15 µm de longueur), phénomène favorisé par la désoxygénation. En effet, la désoxyhémoglobine S présente une propriété naturelle de polymérisation si elle est en solution concentrée (ce qui est le cas dans les hématies) ; le retour de l'HbS à l'état oxygéné provoque la dissociation des polymères.

Le déterminisme génétique de la drépanocytose

La drépanocytose est une maladie autosomale récessive due à une mutation unique du gène de la bêta globine situé sur le chromosome 11. Il s'agit d'une mutation par substitution : le nucléotide A est remplacé par un nucléotide T en position 17, le 6e codon GAG devient alors GTG.

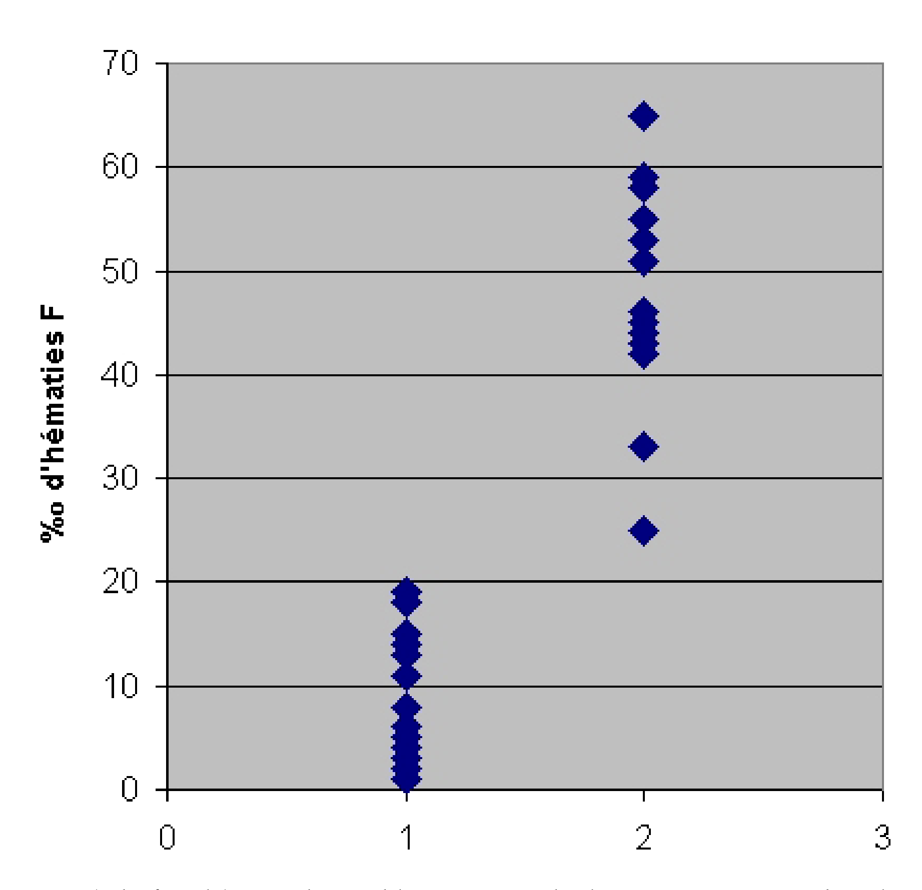
Le gène bêta étant très polymorphe, on connaît plusieurs génotypes drépanocytaires, dont trois prédominent : HbS//HbS (70 %) ; HbS//HbC (25 %), HbS//Hb thalassémie (5 %). Si dans le génotype d'un individu

drépanocytaire, on trouve toujours un allèle HbS, celui-ci peut être associé soit à un autre allèle HbS (génotype HbS//HbS), soit à l'allèle HbC, soit à un allèle Hb thalassémique. L'allèle HbC diffère de l'allèle HbA par une mutation au 6e codon, qui a pour conséquence la substitution de l'acide glutamique par la lysine. Généralement le génotype HbS//HbC se traduit au niveau clinique par un phénotype drépanocytaire nettement moins accusé que celui entraîné par le génotype HbS//HbS. Un allèle Hb thalassémique entraîne un arrêt anticipé de la chaîne bêta au cours de la traduction et donc la synthèse d'une chaîne bêta raccourcie, non fonctionnelle.

La variabilité des phénotypes drépanocytaires

Parmi une population d'individus de génotype HbS//HbS, on a identifié deux catégories de patients : certains développent des crises drépanocytaires graves et fréquentes ; d'autres ne développent qu'exceptionnellement des crises qui restent bénignes.

Les patients qui souffrent de crises caractéristiques de la drépanocytose ont un taux d'hémoglobine F inférieur à 10 ‰ ; ceux qui ne développent que de rares crises ont un taux d'hémoglobine F qui atteint 2 à 7 %.



L'hémoglobine F (Hb fœtale) est détectable à partir de la 5e semaine et c'est le principal constituant hémoglobinique de la vie fœtale. Cette HbF est constituée de deux chaînes α (codées par le gène alpha) et de deux chaînes γ (codées par le gène gamma). Normalement, chez l'adulte, l'HbF ne subsiste qu'à l'état de traces inférieures à 1 %, et reste limitée à une population cellulaire restreinte, les « cellules F » ; ces dernières représentent environ 1 à 7 % des érythrocytes.

Chez un individu produisant de l'HBS, mais qui continue aussi à produire beaucoup d'HBF, les crises drépanocytaires sont rares et peu importantes car l'HBF ne s'intègre pas dans le polymère « poly HBS », ce qui limite la formation de fibres d'HBS et donc la déformation des hématies.

Ainsi, même si le phénotype drépanocytaire est monogénique, il est influencé par l'activité de plusieurs autres gènes, dont le gène gamma.

Pistes d'exploitation pédagogique des données fournies

L'étude du phénotype drépanocytaire avec les données fournies permet d'aborder les notions suivantes :

- le phénotype peut se définir à différentes échelles ;
- le génotype détermine le phénotype moléculaire, lui-même responsable du phénotype cellulaire et clinique ;
- la complexité des relations génotype/phénotype : influence de l'environnement, influence d'autres gènes ;
- un même phénotype peut correspondre à plusieurs génotypes.

Il est particulièrement intéressant ici de coupler l'utilisation du logiciel *RasTop* avec celle du logiciel *Anagène*.

Le phénotype peut se définir à plusieurs échelles

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype drépanocytaire/Drépanocytose** aboutit à **:**

- Allèles Globine Bêta qui charge le fichier alleles-Glob-beta.edi contenant les séquences strictement codantes des allèles bêta A et bêta S du gène de la globine bêta ;

- Protéines Bêta qui charge le fichier prot-glob-beta.edi contenant les séquences protéiques des globines bêta A et bêta S.

Documents fournis

Dans la banque de documents, le chemin **Relations génotype – phénotype à différents niveaux d'organisation du vivant/Le phénotype drépanocytaire** permet d'accéder aux fichiers :

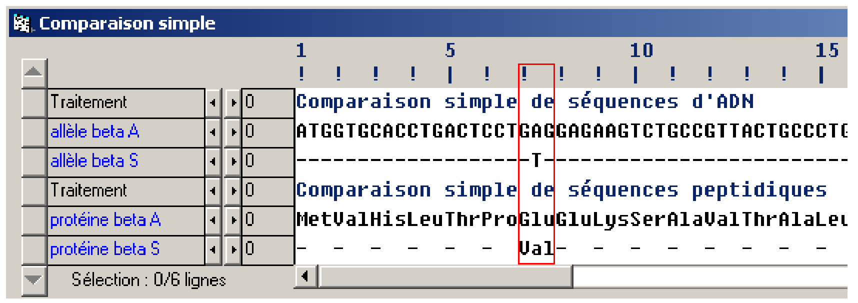
- phenodrep.bmp : texte de présentation du phénotype drépanocytaire au niveau clinique, cellulaire et moléculaire ;
- hematies.bmp : photos d'hématies normales et d'hématies drépanocytaires ;
- brin.bmp : schémas illustrant la formation de brins et de fibres d'HBS à partir de dimères HBS/HBS ;
- dimereHBS.bmp : visualisation 3D obtenue avec le logiciel *RasTop* (fichier hbshbs.pdb) d'un dimère HBS/HBS avec mise en évidence de la valine 6 ;
- comparglobines.bmp : visualisation 3D obtenue avec le logiciel *RasTop* (fichiers globdrep.pdb et globnorm.pdb) d'une globine bêtaA (normale) et d'une globine bêtaS (drépanocytaire).

Fichiers des molécules en 3D

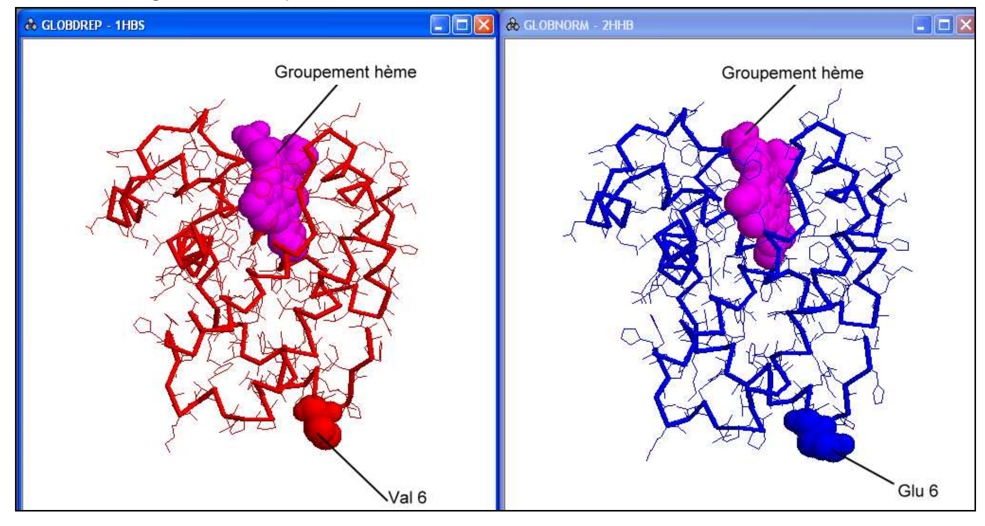
Dans le dossier 3D sous **Anagene2**, le répertoire **Drepanocytose** contient les fichiers .pdb suivants : betanorm.pdb, betadrep.pdb, hbnorm.pdb et hbshbs.pdb.

	Phénotype macroscopique	Phénotype cellulaire	Phénotype moléculaire
Individu sain	Pas de troubles particuliers	Hématies non déformées	Hémoglobine constituée principalement de deux chaînes alpha et deux chaînes **bêtaA**
Individu drépanocytaire	Anémie, fatigue, crises drépanocytaires	Hématies en faucille, déformées par des fibres d'HbS	Hémoglobine constituée principalement de deux chaînes alpha et deux chaînes **bêtaS**

Le génotype détermine le phénotype moléculaire, lui-même responsable du phénotype cellulaire et clinique : la mise en parallèle des comparaisons effectuées entre les deux allèles bêtaA et bêtaS et les protéines correspondantes met en évidence une seule différence : une substitution (A → T au 20e nucléotide) qui modifie le 7e codon, provoquant ainsi le remplacement de l'acide glutamique par la valine.



Cette différence d'un seul acide aminé ne modifie pas la forme de la globine, comme le montre le document ci-dessous, obtenu avec le logiciel *RasTop* :



Globine bêta S *Globine Bêta A*

Attention, dans les séquences fournies pour *Anagène*, le codon d'initiation et donc la méthionine initiale, sont présents ; par contre, dans la molécule accessible en visualisation 3D, cette méthionine initiale n'est pas présente, puisqu'elle a été excisée lors de maturation de la protéine. Cela explique la différence de numérotation entre *Anagène* et *RasTop* : dans le premier cas, c'est l'acide aminé n° 7 qui est modifié, alors que dans de deuxième, c'est le n° 6.

Le remplacement de l'acide glutamique par la valine modifie les propriétés de la globine bêta : la valine forme un point de collage qui favorise la formation de dimères HBS/HBS, et donc la formation de brins et de fibres d'hémoglobine qui vont déformer les hématies ; la circulation capillaire se fait alors plus difficilement, et explique les troubles observés chez les malades.

La complexité des relations génotype/phénotype

Influence de l'environnement

Documents fournis

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes drépanocytaires/Influence de l'environnement** permet d'accéder aux fichiers :

- prevention-drepanocytose.bmp : texte précisant quelques mesures de prévention que doivent respecter les individus drépanocytaires ;
- baisseoxygene.bmp : texte précisant la relation entre une baisse du taux d'oxygène sanguin et l'apparition d'une crise drépanocytaire.

Le phénotype d'un individu de génotype HbS//HbS est influencé par l'environnement : des mesures de prévention permettent de limiter les crises drépanocytaires en évitant les baisses d'oxygénation du sang.

Influence d'autres gènes

Séquences et documents

Fichiers des séquences

Dans la banque des thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes drépanocytaires/Drépanocytoses** aboutit à :

- **Références gène bêta globine** pour accéder à **Allèles gène globine bêta** qui charge le fichier allref-glob-beta.edi affichant quelques allèles du gène de la bêta globine présents dans les différentes familles étudiées – séquences strictement codantes des allèles bêta S (S), bêta A (A) – bêta Tha2 (Tha2) et bêta C (C) ;
- **Dominance et récessivité (Famille 1)** pour accéder à **Allèles famille1** qui charge le fichier allFam1drep.edi affichant des séquences strictement codantes des allèles du gène de la bêta globine chez les individus de la famille 1 [I1, I2 et II1 : (A//S) ; II2 : (S//S)] ;
- **Influence de l'hémoglobine fœtale (Famille 2)** pour accéder à **Allèles famille2** qui charge le fichier allFam2drep.edi affichant des séquences strictement codantes des allèles du gène de la bêta globine chez les individus de la famille 2 [I1 et I2 : (A//S) ; II1 : (S//S) ; II2 : (A//A)]

Documents fournis

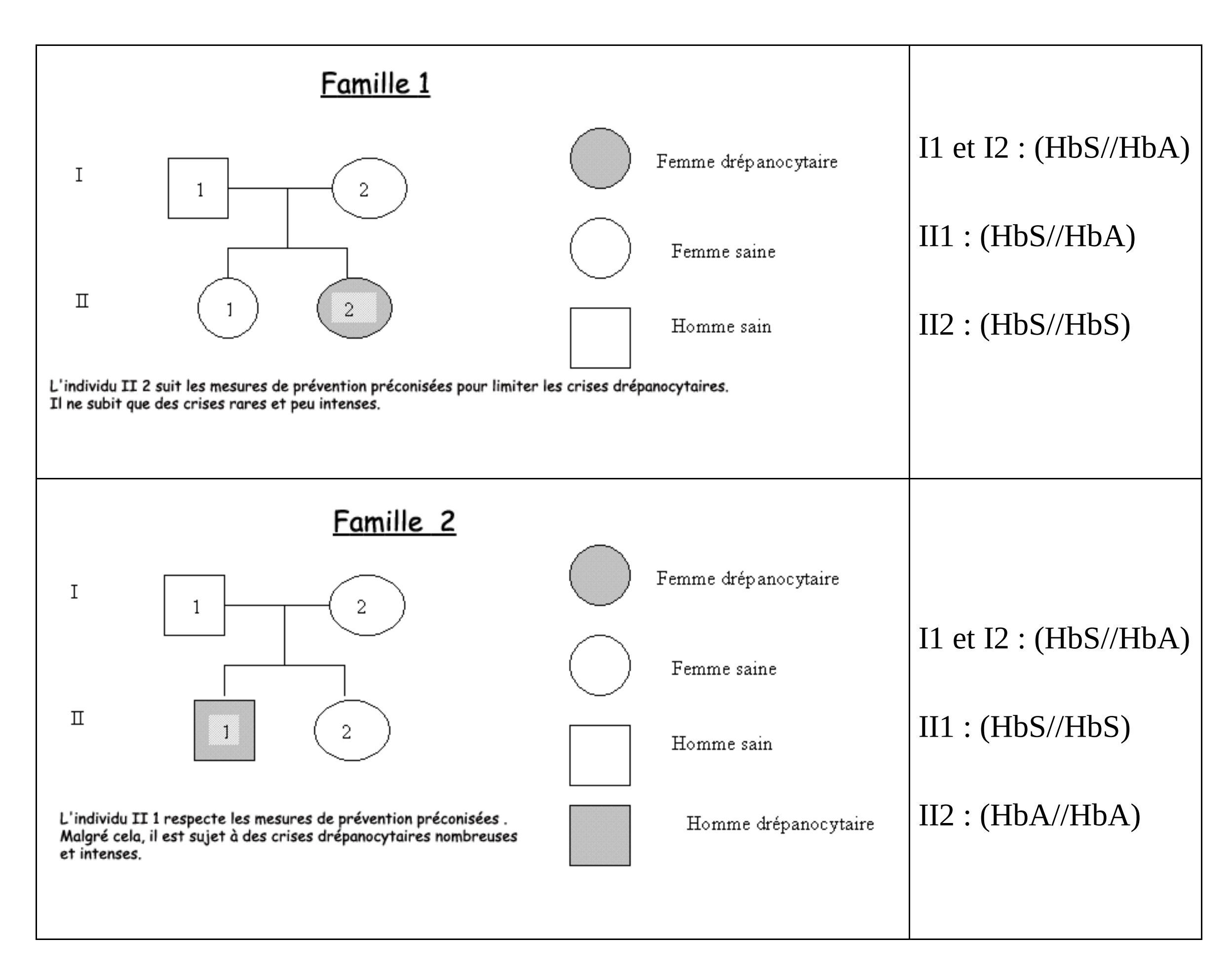
Dans la banque de documents, le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype drépanocytaire** permet d'accéder aux fichiers :

- autresalleles.bmp : texte présentant le polymorphisme du gène de la bêta globine, et présentant notamment les allèles HBC et THA2 ;
- hbfoetale.bmp : texte présentant des résultats de mesures de taux d'hémoglobine fœtale chez des patients drépanocytaires, et permettant de mettre en évidence une atténuation très nette des symptômes drépanocytaires chez les malades produisant beaucoup d'hémoglobine fœtale.

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes drépanocytaires** permet d'accéder à :

- **Dominance et récessivité (Famille 1)** pour charger le fichier arbrefamille1-Drep.bmp qui présente l'arbre généalogique de la famille 1 ;
- **Influence de l'hémoglobine fœtale (Famille 2)** pour charger le fichier arbrefamille2-Drep.bmp qui présente l'arbre généalogique de la famille 2.

La comparaison des allèles du gène bêta de chaque membre de la famille avec les allèles de référence permet de déterminer le génotype de chacun :



Les indications fournies sur les arbres mentionnent que les individus malades des deux familles vivent dans les mêmes conditions et respectent les mêmes mesures de prévention.

La prise en compte des informations apportées par le document concernant l'hémoglobine fœtale permet de conclure à l'influence de l'expression d'autres gènes sur le phénotype drépanocytaire. Les individus de génotype HbS//HbS capables de produire encore beaucoup d'hémoglobine fœtale sont beaucoup moins malades que les autres.



Plusieurs génotypes pour un même phénotype

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes drépanocytaires/Drépanocytoses** aboutit à :

- **Références gène bêta globine** pour accéder à **Allèles gène globine bêta** qui charge le fichier allref-glob-beta.edi affichant les allèles de référence ;
- **Dominance et récessivité (Famille 1)** pour accéder à **Allèles famille1** qui charge le fichier allFam1drep.edi affichant des séquences strictement codantes des allèles du gène de la bêta globine chez les individus de la famille 1 ;
- **Plusieurs génotypes pour un même phénotype (Famille 3)** pour accéder à **Allèles famille3** qui charge le fichier allFam3drep.edi affichant des séquences strictement codantes des allèles du gène de la bêta globine chez les individus de la famille 3 ;
- **Plusieurs génotypes pour un même phénotype (Famille 4)** pour accéder à **Allèles famille4** qui charge le fichier allFam4drep.edi affichant des séquences strictement codantes des allèles du gène de la bêta globine chez les individus de la famille 4.

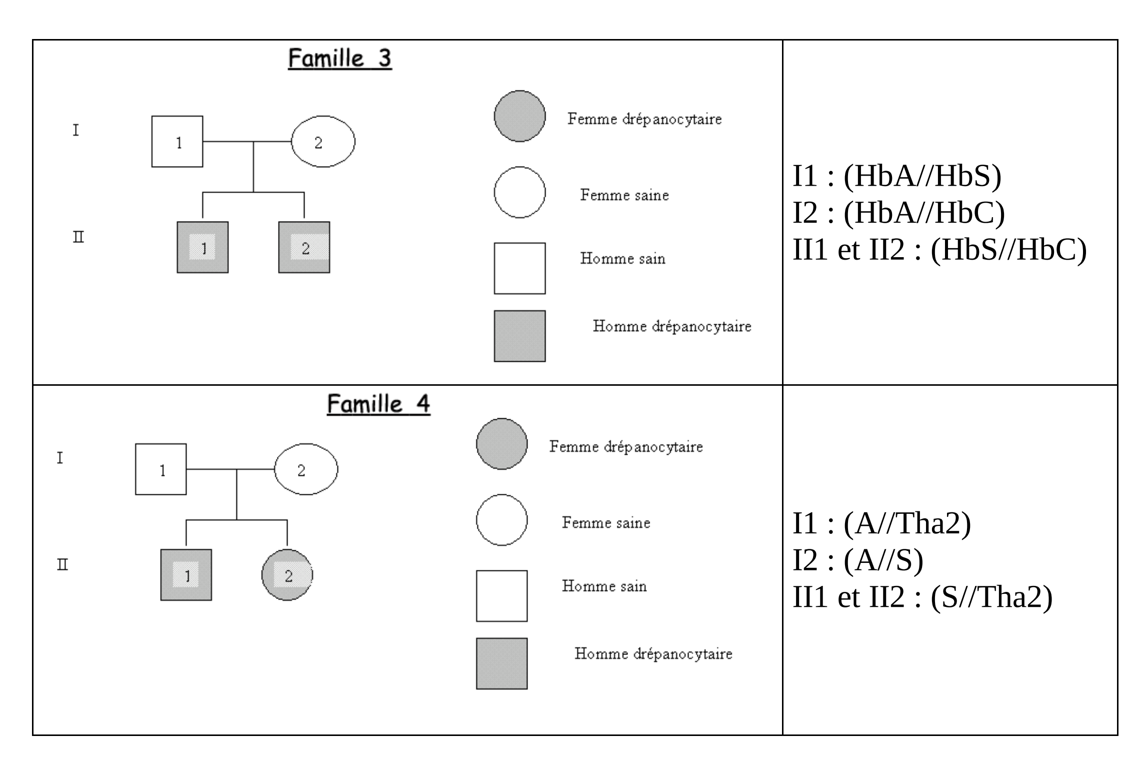
Documents fournis

Dans la banque de documents, le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype drépanocytaire** permet d'accéder au fichier autresalleles.bmp qui affiche un texte présentant le polymorphisme du gène de la bêta globine, et notamment les allèles HBC et THA2.

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes drépanocytaires** aboutit à :

- **Dominance et récessivité (Famille 1)** pour charger le fichier arbrefamille1-Drep.bmp qui présente l'arbre généalogique de la famille 1 ;
- **Plusieurs génotypes pour un même phénotype (Famille 3)** pour charger le fichier arbrefamille3-Drep.bmp qui présente l'arbre généalogique de la famille 3 ;
- **Plusieurs génotypes pour un même phénotype (Famille 4)** pour charger le fichier arbrefamille4-Drep.bmp qui présente l'arbre généalogique de la famille 4.

La comparaison des allèles du gène bêta de chaque membre de la famille avec les allèles de référence permet de déterminer le génotype de chacun des membres de chaque famille :



Les individus II2 de la famille 1, II 1 et 2 des familles 3 et 4 ont le même phénotype clinique : ils sont tous atteints de drépanocytose. Pourtant, leurs génotypes diffèrent.



Les phénotypes phénylcétonuriques

Informations scientifiques

Les caractéristiques d'un phénotype phénylcétonurique – Phénotypes clinique et biochimique

La phénylcétonurie est toujours caractérisée par une augmentation de la concentration plasmatique en phénylalanine (hyperphénylalaninémie).

Le test de Guthrie permet de dépister une hyperphénylalaninémie à la naissance. Ce test consiste à déposer un papier-filtre imbibé d'une goutte de sang du nourrisson sur une culture de *Bacillus* dont la croissance est inhibée par manque de phénylalanine ; selon l'intensité de la croissance bactérienne, on estimera s'il y a hyperphénylalaninémie, qui sera alors confirmée par dosage sanguin.

La phénylalanine est un acide aminé indispensable apporté par l'alimentation, et les besoins quotidiens sont de 200 à 500 mg. Sa concentration physiologique moyenne est de l'ordre de 60 µmoles/l de plasma. Avec une alimentation courante, environ un quart de la phénylalanine est incorporée dans les protéines et le reste est métabolisé par les cellules hépatiques : sous l'action d'une enzyme hépatique, la PAH (phénylalanine hydroxylase), il est métabolisé en tyrosine, adrénaline et mélanine.

L'excès de phénylalanine étant toxique, son accumulation dans le plasma et dans le milieu intracérébral entraîne une arriération psychique et des troubles caractériels, le retard mental étant très sévère si le malade n'est pas pris en charge dès la naissance. Le poids du cerveau des individus phénylcétonuriques est au-dessous de la normale et la myélinisation des fibres nerveuses est défectueuse. L'espérance de vie des malades non traités est considérablement raccourcie : mortalité de 50 % avant 20 ans et de 75 % avant 30 ans.

Le métabolisme de la phénylalanine

$$C_6H_5-CH_2-CH(COO^-)-NH_3^+ + O_2 + \text{Tétrahydrobioptérine}$$

Phénylalanine

↓ Phénylalanine hydroxylase

$$HO-C_6H_4-CH_2-CH(COO^-)-NH_3^+ + H_2O + \text{Quinonoïde dihydrobioptérine}$$

Tyrosine

On parle d'hyperphénylalaninémie lorsque la concentration plasmatique en phénylalanine est supérieure à 120 µmoles/l de plasma, mais les hyperphénylalaninémies entraînant un phénotype phénylcétonurique sont supérieures à 1 000 µmoles/l de plasma.

En cas d'hyperphénylalaninémie supérieure à 1 000 µmoles/l de plasma, une voie parallèle d'élimination urinaire se crée : la phénylalanine est alors décarboxylée et éliminée sous forme d'acide phénylpyruvique.

Les phénotypes moléculaires : l'enzyme PAH

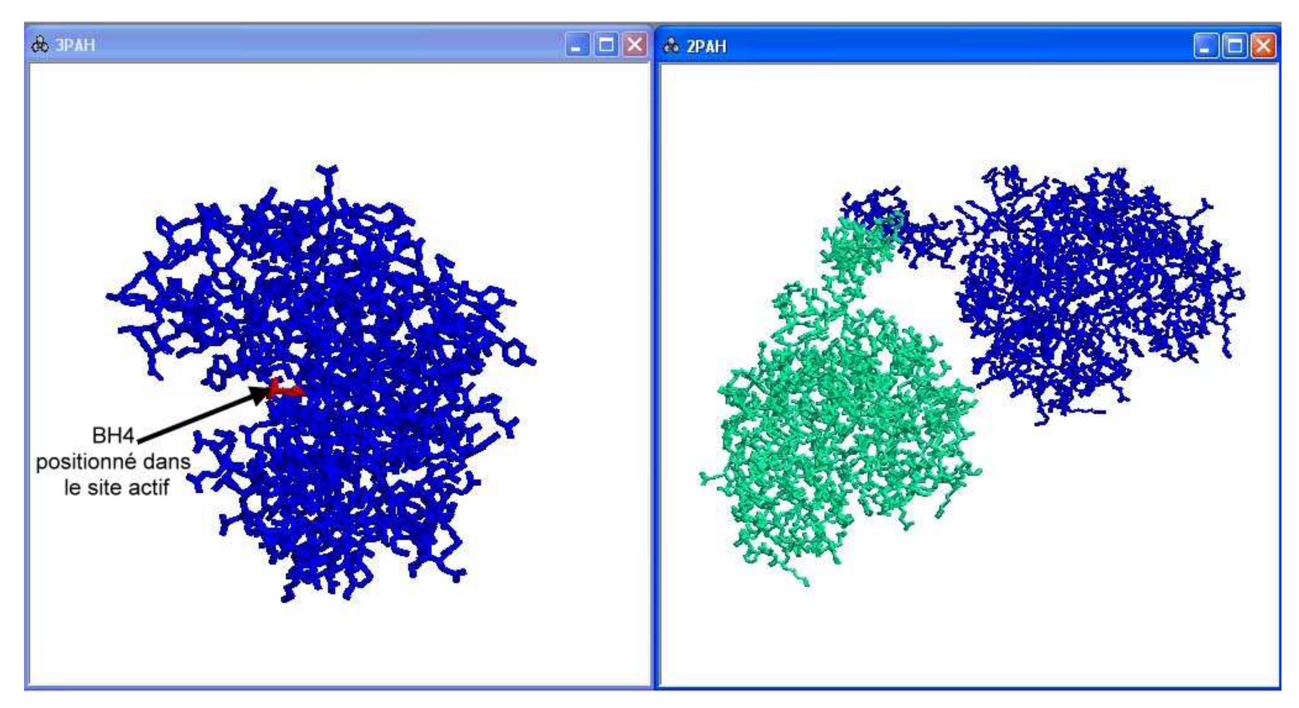
La PAH comprend une partie protéique codée par le gène de la PAH et un cofacteur indispensable à la réaction d'hydroxylation : la tétrahydrobioptérine (ou BH4).

La partie protéique de la PAH

La PAH est une protéine tétramérique.

Le site catalytique met en jeu un atome de fer, trois molécules d'eau et trois résidus : Glu 330, His290 et His285. Ce site catalytique peut être repéré par des modèles montrant l'enzyme associée au cofacteur BH4.

Le document ci-dessous présente un monomère de la PAH et un dimère dans lequel les deux monomères sont associés par l'intermédiaire des hélices α du bout terminal.



Monomère de la PAH *Dimère de la PAH*

Le cofacteur BH4

Le BH4 est synthétisé de novo à partir de guanosine triphosphate (GTP) au cours de quatre réactions enzymatiques catalysées par trois enzymes différentes : GTPCK, PTPS et SR. Il est également recyclé et sa régénération se fait en deux étapes catalysées par les enzymes PCD et DHPR.

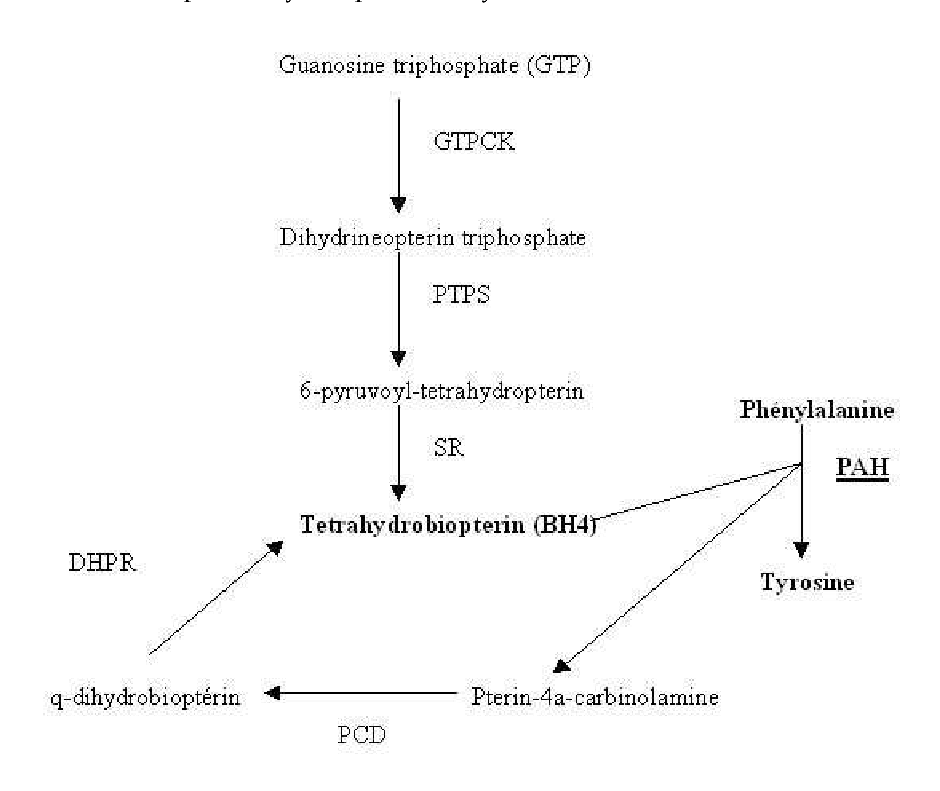


Schéma de la synthèse du BH4

Déterminisme génétique des phénylcétonuries

Les phénotypes phénylcétonuriques peuvent être dus à des mutations dans différents gènes et donc avoir des origines variées.

Implication du gène de la PAH

Le gène codant pour l'enzyme PAH est très polymorphe. Bien que l'on connaisse plusieurs allèles codant pour une PAH fonctionnelle, la plupart des allèles codent pour une enzyme dont l'activité est plus ou moins réduite, voire nulle. Tous ces allèles codant pour une PAH de moindre activité sont récessifs, et la phénylcétonurie ne peut donc apparaître que chez des individus homozygotes.

La comparaison des allèles mutés, de la localisation des mutations et des effets au niveau clinique (voir tableau dans la partie pédagogique) permet de tirer les conclusions suivantes :

- toutes les mutations touchant des acides aminés du site actif entraînent généralement une hyperphénylalaninémie forte ;
- les mutations faux-sens touchant des domaines autres que le site actif ont des conséquences phénotypiques variables selon la position de l'acide aminé modifié ; les mutations touchant les acides aminés de la surface de l'enzyme sont les moins sévères ;
- il est possible d'estimer l'activité de la PAH selon les allèles présents chez les individus homozygotes et hétérozygotes. D'un point de vue clinique, lorsque l'activité est inférieure ou égale à 5 % la phénylcétonurie est sévère ; au-delà de 15 %, la maladie est très atténuée.

	ALLÈLE 1				
ALLÈLE 2	PheMut243	PheMut280	PheMut158	PheMut261	PheMut414
PheMut243	0	1,5	5	15	25
PheMut280	1,5	3	6,5	16,5	26,5
PheMut158	5	6,5	10	20	30
PheMut261	15	16,5	20	30	40
PheMut414	25	26,5	30	40	50

% d'activité de la PAH en fonction du génotype

Implication d'autres gènes

Gènes codant pour des enzymes impliquées dans les réactions de synthèse/recyclage du BH4

Le BH4 étant en quantité limitée dans la cellule, il doit être sans cesse produit et régénéré. Une mutation dans l'un des gènes codant pour une enzyme indispensable à l'une des réactions nécessaires peut donc aussi entraîner un phénotype phénylcétonurique.

On connaît ainsi plusieurs allèles du gène de la DHPR impliqués dans un phénotype phénylcétonurique :

		Séquence nucléique		**Polypeptide**	
Noms des allèles	**Type de mutation**	**Nucléotides changés**	**Codons changés**	**Acides aminés changés**	**Hyperphénylalaninémie de l'homozygote**
DHPRRef		référence	référence	référence	Non
DHPRMut97	délétion non décalante	CC290	TCC97TAA	Ser97Stop	Forte
DHPRMut108	mutation faux-sens	T322G	TGG108GGG	Trp108Gly	Forte
DHPRMut124	insertion non décalante		ACT366-367-368	124Thr	Forte

Autres gènes

Des observations cliniques ont prouvé que des personnes pouvaient avoir des phénotypes phénylcétonuriques nettement différents alors qu'ils avaient le même génotype pour le gène de la PAH et pour les gènes en relation avec le métabolisme du BH4.

Parmi les gènes en cause encore non identifiés se trouvent probablement ceux qui codent pour des protéines assurant le transport des acides aminés à travers la paroi des capillaires sanguins. Des données récentes montrent que pour une même concentration plasmatique en phénylalanine, il peut y avoir des concentrations très variables au sein du tissu cérébral.

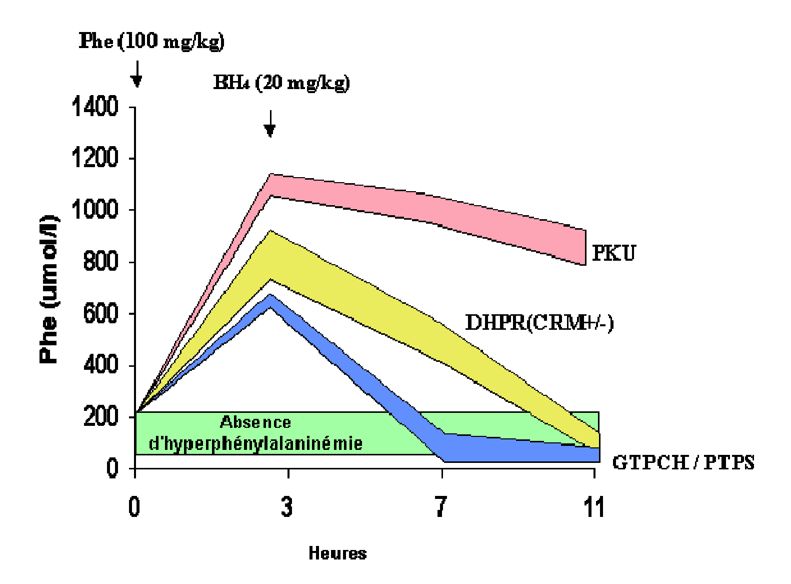
Diagnose différentielle des hyperphénylalaninémies

Un test permet de déterminer l'origine d'une hyperphénylalaninémie. Ce test consiste à faire ingérer des tablettes de phénylalanine à raison de 100 mg/kg, et trois heures plus tard de la BH4 dissoute dans un jus d'orange ou de l'eau. Des analyses sanguines sont réalisées avant l'ingestion de phénylalanine, puis au bout de trois heures, sept heures et onze heures.

L'hyperphénylalaninémie est nette au bout de trois heures ; l'hyperphénylalaninémie modérée au départ s'explique par le fait que les patients suivaient un régime alimentaire depuis leur naissance, le test de Guthrie ayant permis de dépister précocement leur maladie.

En cas de PAH non fonctionnelle, la prise de BH4 n'a aucun effet. En cas de BH4 non recyclé, la prise de BH4 est suivie d'un retour à une phénylalaninémie physiologique normale.

On peut ainsi déterminer si l'hyperphénylalaninémie est due à une mutation touchant le gène de la PAH ou à une mutation touchant l'un des gènes impliqués dans la synthèse ou le recyclage du BH4.



PKU : cas d'un individu pour lequel le phénotype phénylcétonurique est dû à une anomalie de fonctionnement de la PAH ;

DHPR et GTPCH/PTPS : cas d'individus pour lesquels le phénotype phénylcétonurique est dû à une anomalie de recyclage (pour le DHPR) ou de synthèse (pour le GTPCH/PTPS) du BH4 (voir le schéma de la synthèse).

Des informations complémentaires et des animations grâce au logiciel *Chime* sont disponibles sur le site de l'INRP à l'adresse http://www.inrp.fr/Acces/biotic/ et sur le site du BH4.

Interactions avec l'environnement

Si une hyperphénylcétonurie est détectée (concentration en phénylalanine supérieure à 20 mg/dl), un régime alimentaire très strict est mis en place. Ce régime permettra de maintenir le taux de phénylalanine dans le sang à une valeur inférieure au seuil de toxicité.

Régime alimentaire préconisé pour les individus phénylcétonuriques :

- interdiction de consommer : viandes, laitages, poissons, œufs, céréales ;
- possibilité de consommer des fruits et légumes, mais de façon contrôlée ;
- consommation libre d'hydrates de carbones (sucre et amidon), d'huiles et de margarines sans protéines ;
- utilisation de produits diététiques spéciaux pour équilibrer le régime, notamment de produits apportant un mélange d'acides aminés essentiels.

Ce régime devra être poursuivi toute la vie, toute interruption ayant des effets délétères sur les performances intellectuelles des patients.

Les contrôles sanguins du taux de phénylalanine doivent être fréquents de manière à adapter le régime en fonction de changements éventuels.

Pistes d'exploitation pédagogique des données fournies

L'analyse des phénotypes phénylcétonuriques à l'aide des données fournies présente plusieurs intérêts pédagogiques :

- le gène de la PAH est très polymorphe et les enzymes codées par les différents allèles ont des activités différentes, rendant ainsi compte de la diversité des phénotypes phénylcétonuriques, et permettant de discuter des conséquences des mutations sur le phénotype à différents niveaux (moléculaires, biochimique, clinique) ;
- l'enzyme PAH n'a une activité catalytique qu'en présence d'un cofacteur (BH4) dont la synthèse et le recyclage nécessitent l'intervention de plusieurs autres enzymes, codées chacune par un gène. On illustre donc bien ici la complexité des relations génotype/phénotype, plusieurs gènes intervenant dans la réalisation d'un même phénotype ;
- les troubles de la maladie peuvent être supprimés par un régime alimentaire strict et précoce. Là encore la complexité des relations génotype/phénotype est illustrée, un facteur de l'environnement pouvant modifier le phénotype clinique.



La diversité des phénotypes et les relations génotype/phénotype (chez les homozygotes)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype phénylcétonurique** permet grâce au développement de l'arborescence **Phénylcétonurie** d'accéder à :

- **Allèles PAH** qui charge le fichier choixAdnPAH.edi affichant des séquences nucléiques strictement codantes des allèles PheNorm, PheMut39, PheMut55, PheMut95, PheMut111, PheMut222 et PheMut245 ;

- **Protéines PAH** qui charge le fichier choixProtPAH.edi affichant les séquences protéiques des enzymes PAH Norm (activité normale), PheMut39, PheMut55, PheMut95, PheMut111, PheMut222 et PheMut245 (activité réduite) ;

Le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes phénylcétonuriques** permet grâce au développement de l'arborescence **Phénylcétonuries PAH–DHPR** d'accéder à :

- **Références gène PAH/Nucléiques et protéiques/Références PAH ADN** pour charger le fichier RefPAH-ADN.edi affichant des séquences strictement codantes de trois allèles de la PAH (allèle PheNorm codant pour une enzyme PAH fonctionnelle ; allèles Mut95 et Mut111 codant pour des enzymes PAH non fonctionnelles). Ces allèles ont été choisis car ce sont ceux qui sont présents dans les familles étudiées ;

- **Polymorphisme PAH/Allèles PAH** pour charger le fichier PolyPAH.edi affichant les séquences strictement codantes de 31 allèles du gène PAH ; l'allèle PheNorm code pour une enzyme PAH fonctionnelle ; tous les autres allèles codent pour une enzyme PAH non fonctionnelle, et le numéro présent dans leur nom indique la position du codon muté.

Documents fournis

Dans la banque des documents, le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype phénylcétonurique** permet d'accéder aux fichiers :

- reactionPAH.jpg : réaction de transformation de la phénylalanine en tyrosine catalysée par la PAH ;

- phenylcetonurie.bmp : texte de présentation du phénotype phénylcétonurique à différentes échelles ;

- monomerePAH.jpg : image obtenue à l'aide du logiciel *RasTop*, présentant un monomère de la PAH et localisant le site actif la position du cofacteur BH4 ;

- effetMutPAH.bmp : tableau présentant le phénotype clinique des individus mutants homozygotes.

Fichiers des molécules en 3D

Dans le dossier 3D sous **Anagene2**, le répertoire **Phenylcetonurie** contient les fichiers .pdb suivants : Dimere4PAH.pdb et Monomere2PAH.pdb.

Les documents phenylcetonurie.bmp, reactionPAH.jpg et monomerePAH.jpg présentent le phénotype phénylcétonurique et permettent de le définir à différentes échelles :

- phénotype clinique : symptômes de la maladie ;
- phénotype biochimique : concentration plasmatique en phénylalanine ;
- phénotype moléculaire : enzyme PAH.

La comparaison des séquences (protPAH.edi) des différentes protéines PAH rencontrées chez des individus atteints (enzymes non fonctionnelles) avec la protéine enzymatique de référence PheNorm (enzyme fonctionnelle) et la mise en relation avec les phénotypes cliniques correspondants (effetmutPAH.bmp)

permettent de discuter des relations entre phénotype moléculaire et phénotype clinique. Les différences peuvent être relevées dans un tableau, où l'on précisera aussi le type de mutation (substitution, délétion, addition) et le phénotype clinique correspondant.

Le tableau ci-dessous donne les caractéristiques des allèles présents dans la banque de données fournie, des protéines correspondantes et précise les phénotypes cliniques des homozygotes :

	Séquence nucléique		Polypeptide		
Noms des allèles	Nucléotides changés Nature - position	Codons changés Nature - position	Acides aminés changés Nature - position	Type de mutation	Hyper-phénylalaninémie de l'homozygote
PHENORM	Référence	Référence	Référence	Référence	Non
PHEmut39	T115C	TTC39CTC	Phe39Leu	Substitution faux-sens	Modérée
PHEmut55	-T165	TTT55… TAA 60	codon stop 60	Délétion décalante	Forte
PHEmut95	-ATC283,284,285	-95ATC	-95Ile	Délétion non décalante	Légère
PHEmut104	C311A	GCC104GAC	Ala104Asp	Substitution faux-sens	Légère
PHEmut111	C331T	CGA111TGA	Arg111-	Substitution non-sens	Forte
PHEmut158	G473A	CGG158CAG	Arg158Gln	Substitution faux-sens	Modérée
PHEmut174	T520C	ATC174ACC	Ile174Thr	Substitution faux-sens	Forte
PHEmut176	C526T	CGA176TGA	codon stop 176	Substitution non-sens	Forte
PHEmut187	G561A	TGG187TGA	codon stop 187	Substitution non-sens	Forte
PHEmut194	T581C	CTG194CCG	Leu194Pro	Substitution faux-sens	Forte
PHEmut222	-GA663,664	GAT222TAA	Asp222-	Délétion décalante	Forte
PHEmut232	A696G	CAA232CAG	Gln232Gln	Substitution neutre	Non
Phemut241	G722A	CGC241CAC		Substitution faux-sens	Légère
PHEmut243	C727T	CGA243TGA	Arg243-	Substitution non-sens	Forte
PHEmut245	G735A	GTG245GTA	Val245Val	Substitution neutre	Non
PHEmut261	G782A	CGA261CAA	Arg261Gln	Substitution faux-sens	Légère
PHEmut272	G814T	GGA272TGA	Gly272-	Substitution non-sens	Forte
PHEmut277	T829G	TAT277GAT	Tyr277Asp	Substitution faux-sens	Légère
PHEmut280	G838A	GAA280AAA	Glu280Lys	Substitution faux-sens	Forte
PHEmut281	C842T	CCT281CTT	Pro281Leu	Substitution faux-sens	Forte
PHEmut282	G844A et C846A	GAC282AAA	Asp282Lys	2 substitutions faux-sens	Forte
PHEmut285	C853T	CAT285TAT	His285Tyr	Substitution faux-sens	Forte
PHEmut299	T896G	TTT299TGT	Phe299Cys	Substitution faux-sens	Forte
PHEmut331	T991C	TTT331CTT	Phe331Leu	Substitution faux-sens	Forte
PHEmut346	-G1038	GGG346GGC etc…	Leu347Ser etc…	Délétion décalante	Forte
PHEmut349	T1045C	TCA349CCA	Ser349Pro	Substitution faux-sens	Forte
PHEmut365	-CTC1093,1094,1095	-365CTC	-365Leu	Délétion non décalante	Forte
PHEmut408a	G1223A	CGG408CAG	Arg408Gln	Substitution faux-sens	Légère
PHEmut408b	C1222T	CGG408TGG	Arg408Trp	Substitution faux-sens	Forte
PHEmut414	A1241G	TAC414TGC	Tyr414Cys	Substitution faux-sens	Légère

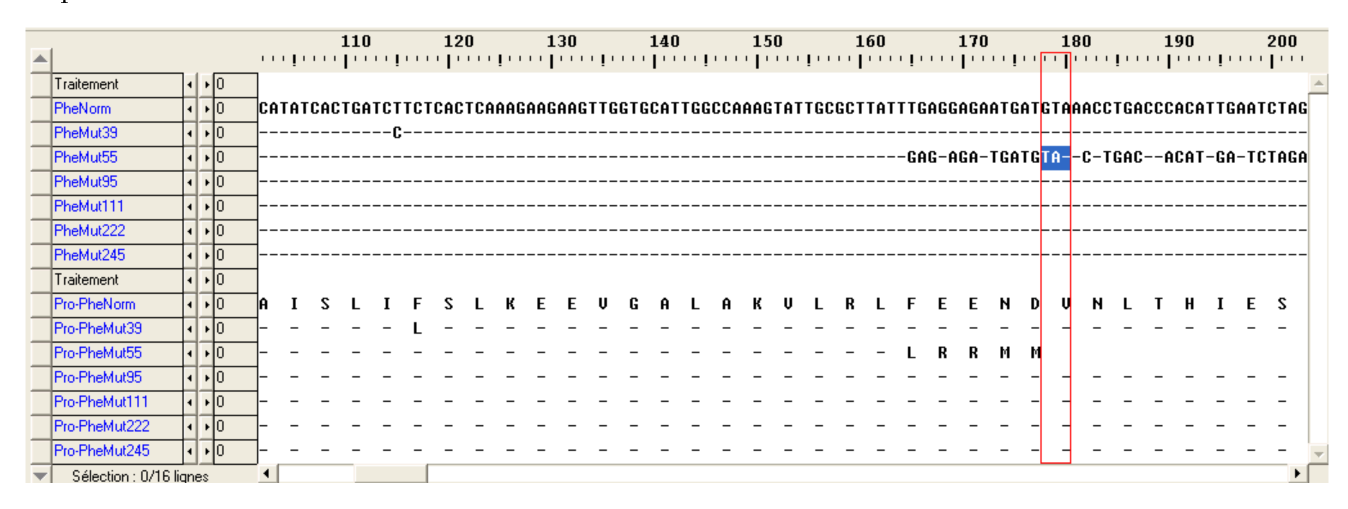
Les connaissances sur les protéines enzymatiques seront mobilisées pour expliquer les observations.

La localisation des mutations sur un monomère de PAH à l'aide du logiciel *RasTop* complète très bien cette activité ; on peut en effet alors distinguer les mutations touchant le site actif des autres mutations et discuter de l'effet de la position des mutations.

On montrera ainsi que les phénotypes alternatifs sont dus à des différences dans les protéines concernées, et qu'il existe plusieurs génotypes possibles pour un même phénotype clinique.

La mise en relation des différences observées au niveau des quelques allèles sélectionnés ici (Allèles PAH et Protéines PAH) avec les différences observées au niveau protéique permet de discuter de l'effet des mutations.

Remarque pratique : bien que les délétions et additions soient plus facilement repérables en utilisant un alignement avec discontinuités, la mise en relation des différences observées entre les allèles avec celles observées entre les polypeptides est facilitée si l'on demande des comparaisons simples dans les deux cas. Ainsi, dans le document ci-dessous, on met facilement en relation pour l'allèle PheMut55 la présence d'un codon stop précoce avec une chaîne polypeptidique plus courte, en relation avec la délétion du nucléotide T en position 165 :



Ces sept séquences ont été retenues car elles illustrent la diversité des mutations et de leurs conséquences (voir tableau précédent), mais il est tout à fait possible de faire travailler les élèves sur les autres allèles : il suffit alors de sélectionner les allèles parmi tous ceux qui sont proposés (polyPAH.edi) et d'en demander la traduction.

Conclusion

Des mutations variées sont à l'origine du polymorphisme du gène de la PAH. Les conséquences de ces mutations sur le phénotype moléculaire sont variables :

- apparition précoce d'un codon stop (mutation non-sens ou mutation décalante), donc chaîne polypeptidique plus courte et non fonctionnelle – PheMut55, 111, 176, 187, 222, 243, 272, 346 ;
- modification importante de la séquence d'acides aminés (délétion d'un ou plusieurs nucléotides) – PheMut95 et 365 ;
- modification ponctuelle de la séquence d'acides aminés (mutation faux-sens) – PheMut39, 158, 194, 241, 261, 277, 280, 281, 282, 285, 299, 331 ;
- aucune modification de la séquence d'acides aminés (mutation neutre) – PheMut232.

Les conséquences des modifications de la séquence d'acides aminés sont variables selon la position des acides aminés modifiés (on ne peut aboutir à cette conclusion que si l'on a complété l'étude sur les séquences par la localisation des mutations sur la molécule en 3D à l'aide d'un logiciel de visualisation moléculaire comme *RasTop* ou *Chime*). Si les acides aminés modifiés sont au niveau du site actif de l'enzyme ou participent à sa structure tertiaire, l'enzyme sera inactive ou très peu active. Si le polypeptide est beaucoup plus court que la normale, l'enzyme ne pourra pas prendre sa conformation spatiale et sera inactive.

Complexité des relations entre gènes, phénotype et environnement

La mise en évidence de la complexité des relations génotype/phénotype peut être abordée à partir de l'exemple des trois familles proposées. Pour chaque famille, sont mises à disposition les données suivantes :

- un arbre qui n'est là que pour préciser le phénotype des membres d'une famille donnée (et non pour initier un raisonnement en fonction des lois de l'hérédité) ;
- les séquences des deux allèles de la PAH que possèdent tous ou certains membres de la famille ;
- les séquences des allèles PAH de référence (allèle PheNorm et les deux allèles mutés présents dans l'une ou l'autre des familles : PheMut95 et PheMut111).

Étude de la famille 1

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes phénylcétonuriques** permet grâce au développement de l'arborescence **Phénylcétonuries PAH-DHPR** d'accéder à :

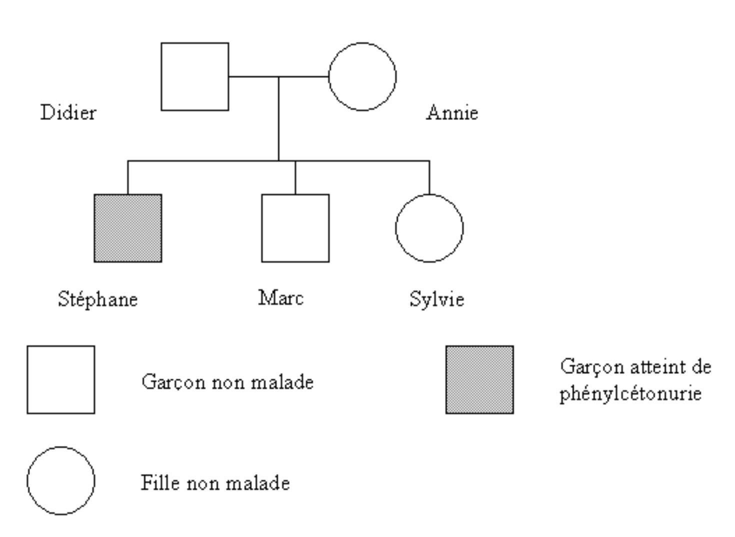
- **Références PAH ADN** qui charge le fichier RefPAH-ADN.edi affichant des séquences strictement codantes de trois allèles de la PAH (allèle PheNorm codant pour une enzyme PAH fonctionnelle ; allèles Mut95 et Mut111 codant pour des enzymes PAH non fonctionnelles). Ces allèles ont été choisis car ce sont ceux qui sont présents dans les familles étudiées ;

- **Famille1/Génotypes famille1 PAH** qui charge le fichier allfam1PAH.edi affichant les séquences strictement codantes des allèles du gène de la PAH de chacun des membres de la famille 1 ; le génotype de Didier, d'Annie et de Marc est PheNorm//PheMut95, celui de Stéphane est PheMut95//PheMut95, et celui de Sylvie est PheNorm//PheNorm.

Documents fournis

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes phénylcétonuriques** permet d'accéder par **Dominance et récessivité (Famille 1)** au fichier fam1phenyl.bmp présentant l'arbre généalogique de la famille 1, avec indication des phénotypes. Dans cette famille, le phénotype phénylcétonurique des individus malades est dû à l'allèle PheMut95 du gène de la PAH.

L'étude de cette famille permet de se familiariser avec la méthode de détermination des génotypes et de discuter des relations de dominance/récessivité.



La comparaison des allèles de chaque individu avec les allèles de référence de la PAH fournis permet de déterminer le génotype de chacun des membres de la famille : le génotype de Didier, d'Annie et de Marc est PheNorm//PheMut95, celui de Stéphane est PheMut95//PheMut95, et celui de Sylvie est PheNorm//PheNorm.

La mise en relation du génotype et du phénotype pour les individus hétérozygotes permet ensuite de discuter des relations de dominance/récessivité (ces relations ne peuvent se définir qu'à l'échelle du phénotype clinique et non à l'échelle moléculaire ; en effet, les deux allèles s'expriment dans la cellule).

Étude de la famille 2

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes phénylcétonuriques** permet grâce au développement de l'arborescence **Phénylcétonuries PAH-DHPR** d'accéder à :

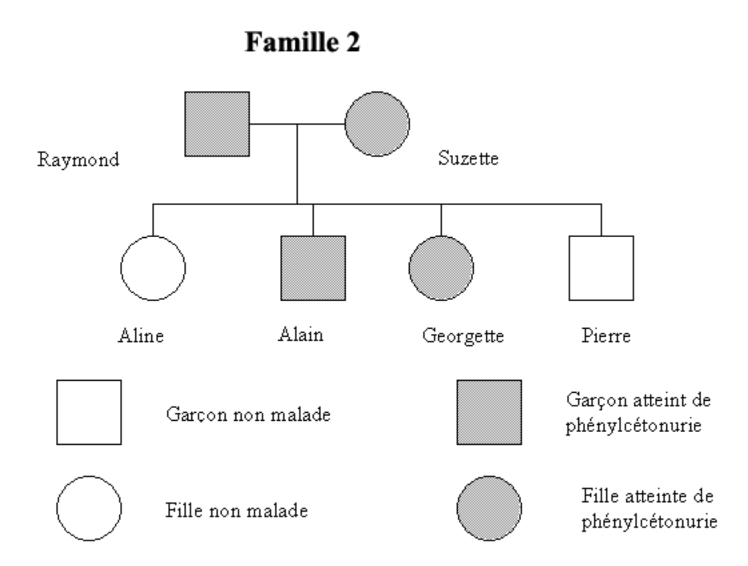
- **Références PAH ADN** qui charge le fichier RefPAH-ADN.edi affichant des séquences strictement codantes de trois allèles de la PAH (allèle PheNorm codant pour une enzyme PAH fonctionnelle ; allèles Mut95 et Mut111 codant pour des enzymes PAH non fonctionnelles). Ces allèles ont été choisis car ce sont ceux qui sont présents dans les familles étudiées ;
- **Génotypes famille2 PAH** qui charge le fichier AllFam2PAH.edi affichant les séquences strictement codantes des allèles du gène de la PAH de chacun des membres de la famille 2. Le génotype de Raymond, Alain et Georgette est PheMut95//PheMut95 ; celui de Suzette, Aline et Pierre est PheNorm//PheMut95 ;
- **Génotypes famille2 DHPR** qui charge le fichier AllFam2DHPR.edi affichant les séquences strictement codantes des allèles du gène de la DHPR de chacun des membres de la famille 2. Le génotype de Raymond, Aline, Georgette et Pierre est DHPRNorm//DHPRMut1 ; celui d'Alain et de Suzette est DHPRMut1//DHPRMut1 ;
- **Allèles DHPR** qui charge le fichier PolyDHPR.edi affichant les séquences strictement codantes de quatre allèles du gène de la DHPR ; l'allèle DHPNorm code pour une enzyme DHPR fonctionnelle ; les allèles DHPRMut1, DHPRMut2 et DHPRMut3 codent pour des enzymes DHPR non fonctionnelles ;
- **Références DHPR Pro** qui charge le fichier protDHPR.edi affichant les séquences protéiques des enzymes DHPR Norm (fonctionnelles), DHPRMut1, DHPRMut2 et DHPRMut3 (non fonctionnelles).

Documents fournis

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes phénylcétonuriques** permet d'accéder par **Plusieurs génotypes pour un même phénotype (Famille 2)** aux fichiers :

- Fam2Phenyl.bmp qui présente l'arbre généalogique de la famille 2, avec indication des phénotypes. Dans cette famille, le phénotype phénylcétonurique des individus malades est dû à l'allèle DHPRMut2 du gène de la DHPR ; tous les individus de cette famille possèdent un phénotype PheNorm//Phenorm ou PheNorm//PheMut95.
- BH4.bmp qui affiche un texte de présentation du BH4 et de son importance et réactions de recyclage et de synthèse.

L'étude de cette famille permet de découvrir un aspect de la complexité des relations génotype/phénotype : un phénotype peut dépendre de l'expression de plusieurs gènes. Il s'agit de mettre l'élève en situation de découverte du problème.



La comparaison des allèles du gène PAH de chaque individu avec les allèles de référence de la PAH permet de déterminer le génotype de chacun : le génotype de Raymond, Alain et Georgette est PheMut95//PheMut95 ; celui de Suzette, Aline et Pierre est PheNorm//PheMut95.

Si l'on s'en tenait au fait que le phénotype phénylcétonurique ne dépende que du seul gène de la PAH, et que l'allèle PheMut95 est récessif par rapport à PheNorm, il y a une incohérence pour Suzette ; possédant un allèle

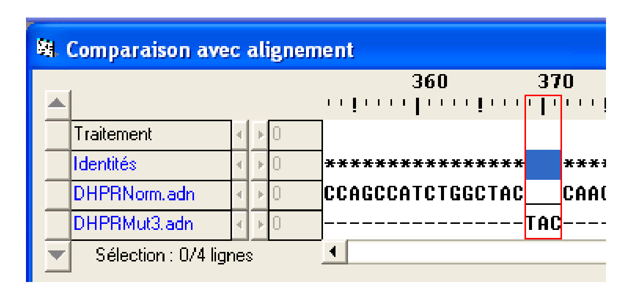
PheNorm, elle ne devrait pas être malade. Il faut donc imaginer que le phénotype phénylcétonurique ne dépend pas que du gène de la PAH.

Des renseignements complémentaires sont alors fournis concernant le fonctionnement de la PAH (BH4.bmp) : nécessité d'un cofacteur, le BH4, qui doit être en permanence recyclé et synthétisé. Cette synthèse/recyclage du BH4 nécessite plusieurs réactions chimiques catalysées chacune par une enzyme différente, codée chacune par un gène différent. Il peut donc y avoir plusieurs origines à la non-transformation de la phénylalanine en tyrosine : inactivité de l'enzyme PAH, mais aussi manque de cofacteur BH4 par déficience d'une des enzymes nécessaires à sa synthèse.

Un travail peut être effectué sur les allèles du gène DHPR et les protéines correspondantes pour discuter, comme il a été fait pour la PAH, des relations entre génotype/phénotype moléculaire/phénotype clinique. Les mutations présentes dans les allèles entraînent des modifications de la séquence d'acides aminés de la protéine DHPR, responsables de l'inactivité de l'enzyme et d'un phénotype phénylcétonurique.

		280 290 300 310 320 330 340 350 360 370
Traitement	0	
DHPRNorm.adn	0	CAATGCCAAATCCAAGTCTCTCTTTAAGAACTGTGACCTGATGTGGAAGCAGAGCATATGGACATCGACCATCTCCAGCCATCTGGCTACCAAGCATCTCAA
DHPRMut1.adn	0	-----------A-GTCTCTCT--A-GA-CTGTGAC-TGATGTG-A-GCAGAGCATATG-ACATCGAC-ATCTC-AGC-ATCTG-CTAC-A-GCATCTCA-G
DHPRMut2.adn	0	---G--
DHPRMut3.adn	0	--T-CCA-G-ATCT
Traitement	0	
Pro-DHPRNorm.adn	0	G N A K S K S L F K N C D L M W K Q S I W T S T I S S H L A T K H L
Pro-DHPRMut1.adn	0	- - - - - S L S L R T V T
Pro-DHPRMut2.adn	0	- - - - - - - - - - - - - - - G - - - - - - - - - - - - - - - - -
Pro-DHPRMut3.adn	0	- T K H
Sélection : 0/10 lignes		

Résultats de comparaisons simples entre les allèles de la DHPR d'une part et entre les protéines DHPR correspondantes d'autre part



Résultat d'un alignement avec discontinuité entre les allèles DHPRNorm et DHPRMut 3

L'allèle DHPRMut1 diffère de l'allèle DHPRNorm par une délétion d'un nucléotide C en position 291, ce qui se traduit par l'apparition d'un codon stop précoce et une protéine plus courte, non fonctionnelle.

L'allèle DHPRMut2 diffère de l'allèle DHPRNorm par une substitution d'un seul nucléotide en position 322 (T → G). Cela se traduit par le changement d'un acide aminé dans la séquence de l'enzyme, ce qui suffit à la rendre non fonctionnelle.

L'allèle DHPRMut3 diffère de l'allèle DHPRNorm par l'addition d'un triplet TAC, ce qui se traduit par l'addition d'un acide aminé dans la séquence de l'enzyme, ce qui suffit à la rendre non fonctionnelle.

La comparaison des allèles du gène DHPR des membres de la famille 2 avec les allèles de référence de ce gène (« allfam2DHPR.edi »-« refDHPR.edi ») permet de déterminer le génotype de chacun pour ce gène : le génotype de Raymond, Aline, Georgette et Pierre est DHPRNorm//DHPRMut1 ; celui d'Alain et de Suzette est DHPRMut1//DHPRMut1.

Une discussion sur le phénotype d'Aline par exemple permet de déterminer les relations de dominance/récessivité entre ces deux allèles. L'étude du génotype de Suzette permet d'expliquer l'incohérence constatée auparavant : Suzette présente un phénotype phénylcétonurique alors qu'elle possède un allèle PAH normal, car elle est homozygote DHPRMut1//DHPRMut1, et ne peut donc synthétiser une enzyme DHPR fonctionnelle. Le recyclage du BH4 ne peut donc se faire et, par conséquent, la cellule manque très rapidement de BH4, donc l'enzyme PAH ne peut remplir son rôle correctement.

Étude de la famille 3

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes phénylcétonuriques** permet grâce au développement de l'arborescence **Phénylcétonuries PAH-DHPR** d'accéder à :

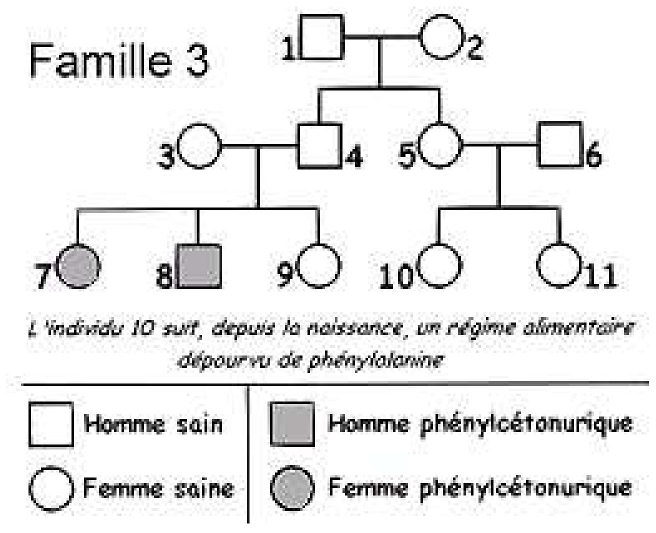
- **Références PAH ADN** qui charge le fichier RefPAH-ADN.edi affichant des séquences strictement codantes de trois allèles de la PAH (allèle PheNorm codant pour une enzyme PAH fonctionnelle ; allèles Mut95 et Mut111 codant pour des enzymes PAH non fonctionnelles). Ces allèles ont été choisis car ce sont ceux qui sont présents dans les familles étudiées ;
- **Génotypes famille3 PAH** qui charge le fichier AllFam3PAH.edi affichant les séquences strictement codantes des allèles du gène de la PAH des individus 7, 8 et 10 de la famille 3. Le génotype de chacun de ces individus est PheMut111/PheMut111.

Documents fournis

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes phénylcétonuriques** permet d'accéder par **Influence de l'environnement (Famille 3)** aux fichiers :

- Fam3Phenyl.bmp qui présente l'arbre généalogique de la famille 1, avec indication des phénotypes. Dans cette famille, trois individus malades ont le même génotype PheMut111//PheMut111, mais l'un d'entre eux n'est pas malade car il suit un régime alimentaire particulier depuis sa naissance ;
- regime.bmp qui affiche un texte présentant le traitement de la phénylcétonurie par un régime alimentaire strict.

L'étude de cette famille permet de découvrir un autre aspect de la complexité des relations génotype/phénotype : l'intervention d'un facteur de l'environnement sur le phénotype.



La comparaison des allèles du gène PAH des individus 7, 8 et 10 avec les allèles de référence permet de déterminer le génotype de ces individus : ils sont tous PheMut111//PheMut111.

Un problème se pose alors car ces trois individus n'ont pas le même phénotype : deux d'entre eux sont malades, mais pas le troisième.

L'arbre généalogique fournit une information complémentaire concernant leur régime alimentaire. L'hypothèse de l'influence de ce régime sur le phénotype peut être confirmée par des renseignements complémentaires concernant ce régime (*cf.* fichier regime.bmp).

Les phénotypes groupes sanguins

Informations scientifiques

La découverte des groupes sanguins

C'est **Richard Lower**, médecin britannique, qui effectua le premier, en 1665, des transfusions sanguines d'animal à animal. Mais le premier à transfuser avec succès du sang d'animal, un agneau, à un jeune garçon anémique de 16 ans, deux ans plus tard, fut **Jean-Baptiste Denis**.

Cependant, les incidents se multiplièrent, et, si la transfusion réussissait parfois, elle ne faisait souvent que précipiter le décès (l'agglutination des hématies du donneur suivie d'une lyse de ces hématies entraînait des troubles métaboliques importants, la destruction des reins…). Bientôt, les transfusions sanguines furent interdites et n'allaient être reprises que vers la fin du siècle dernier.

Lorsque **Karl Landsteiner** (1868-1943) commença ses recherches comme assistant à l'Université viennoise de l'hygiène, il se proposa de surmonter ces difficultés. Il avait constaté que, lorsque l'on mélangeait du sang de plusieurs personnes, on observait parfois une agglutination des hématies et parfois non. Il rassembla alors des échantillons de sang de ses collègues et les mélangea deux par deux. C'est ainsi qu'il découvrit l'existence des groupes sanguins, qu'il nomma A, B et O, et il fit connaître les résultats de ses recherches en 1900. Le quatrième groupe sanguin (AB) ne fut découvert que l'année suivante. Cette découverte permit d'établir une méthode de transfusion sanguine sûre.

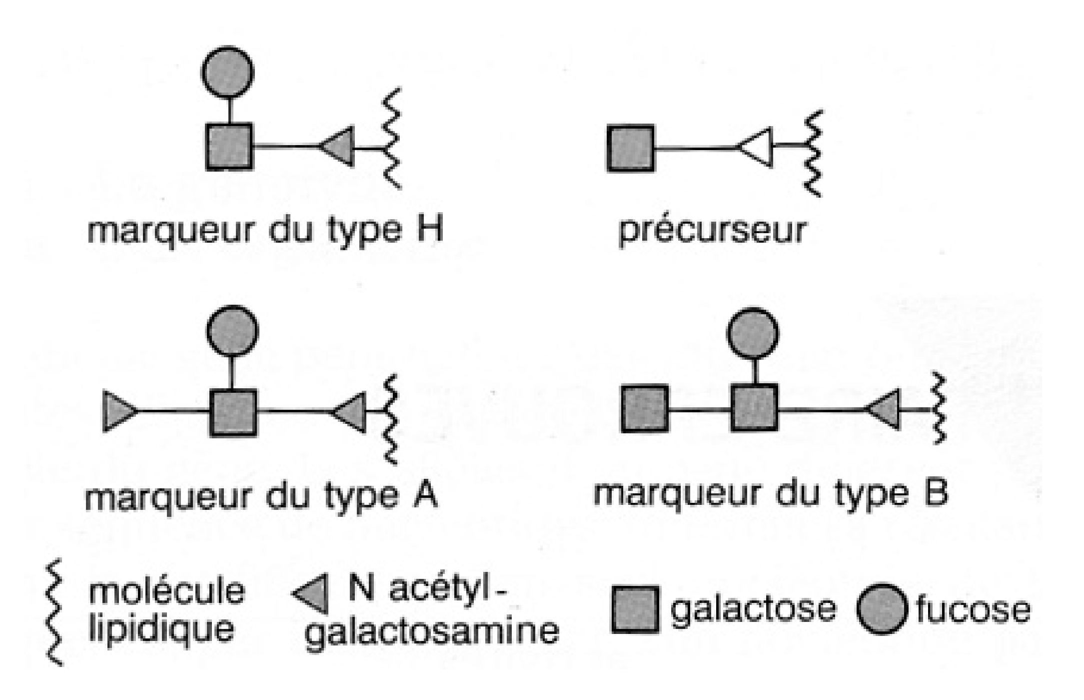
En reconnaissance pour ses travaux, Landsteiner fut nommé professeur de pathologie à l'Université de Vienne, mais il ne s'y plut pas, et partit pour les Pays-Bas puis pour l'Institut Rockfeller de New York. Il découvrit plus tard les groupes sanguins M et N, ainsi que le facteur rhésus, en collaboration avec Philip Levine. Lansdsteiner reçut le prix Nobel de médecine en 1930.

Les marqueurs des groupes sanguins principaux

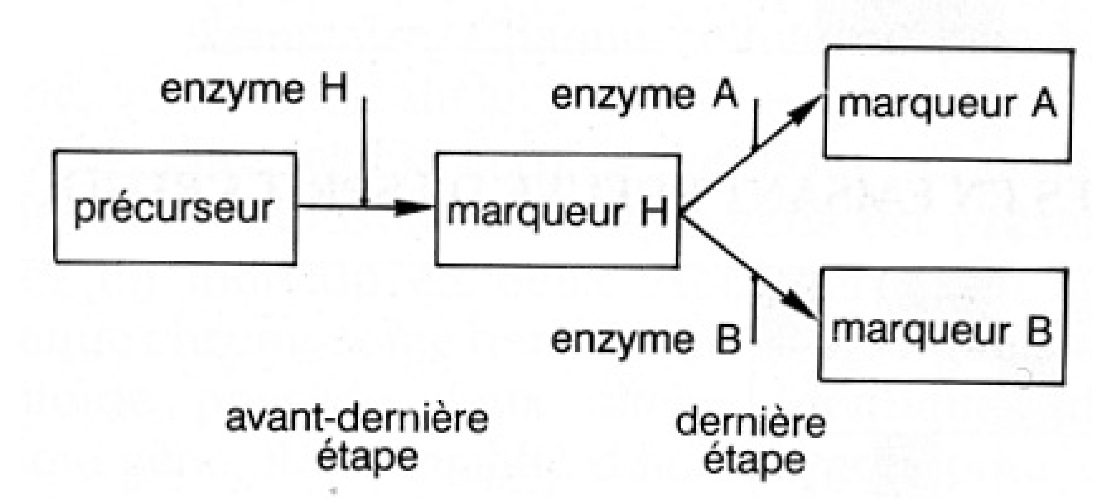
Les marqueurs des groupes sanguins principaux sont des glycolipides présents dans la membrane des hématies. Il existe deux types de marqueurs : des marqueurs de type A et des marqueurs de type B.

Ces marqueurs déterminent les groupes sanguins principaux :

- présence de marqueurs A seulement : groupe sanguin A ;
- présence de marqueurs B seulement : groupe sanguin B ;
- présence de marqueurs A et de marqueurs B : groupe sanguin AB ;
- absence de marqueurs A et B : groupe sanguin O.



Les deux dernières étapes de la chaîne de biosynthèse des marqueurs A et B



Le gène Fut1 code pour la synthèse de l'enzyme H (EH). Le gène des groupes sanguins principaux code pour la synthèse des enzymes A et B (EA, EB).

Les phénotypes sanguins A, B et O

Ils sont définis respectivement par la possession des antigènes membranaires A, B et H. Cependant, il est possible de retrouver les divers niveaux d'expression du phénotype vus avec les hémoglobinopathies :

- phénotype macroscopique : il ne se révèle pas spontanément mais il se manifeste si l'on met les globules rouges d'un individu en contact avec des sérums tests anti A et anti B et aussi en cas de transfusions incompatibles ;
- phénotype cellulaire : c'est la panoplie des marqueurs relatifs à ce système de groupes sanguins présents dans la membrane des hématies mais aussi à la surface des autres cellules de l'organisme ;
- phénotype moléculaire : les marqueurs membranaires résultent de l'action de l'enzyme produit de l'expression du gène ; cette enzyme est en somme le phénotype moléculaire primaire.

Pistes d'exploitation pédagogiques des données fournies

L'analyse des données fournies concernant le phénotype groupes sanguins permet d'aborder les notions suivantes :

- le phénotype peut se définir à différents niveaux, et le phénotype moléculaire détermine le phénotype cellulaire et macroscopique ;
- le phénotype moléculaire s'explique par le génotype ;
- un même phénotype peut correspondre à plusieurs génotypes ;
- plusieurs gènes peuvent être impliqués dans la réalisation d'un phénotype.

La diversité des phénotypes et les relations génotype/phénotype (chez les homozygotes)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype groupes sanguins** permet, par le développement de l'arborescence, d'accéder à :

- **Allèles ABO** qui charge le fichier RefABO-ADN.edi affichant les séquences nucléiques strictement codantes des allèles A101, B101 et O01 ;

- **Protéines ABO** qui charge le fichier refABO-prot.edi affichant les séquences protéiques des enzymes A 101, B101 et O 01 ;

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes groupes sanguins** permet, par le développement de l'arborescence, d'accéder à **Allèles ABO** qui charge le fichier polyABO.edi affichant les séquences nucléiques strictement codantes de dix allèles du gène ABO (allèles A 101, A 102, A 103, A 201, B 101, B 102, B 103, O 01, O 04, O 06) ; tous les allèles A et B fournis codent pour des enzymes A et B fonctionnelles.

Documents fournis

Dans la banque de documents, le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype groupes sanguins** permet d'afficher les fichiers :

- marqueursABO.bmp : représentation schématique des marqueurs membranaires A, B et H ;

- biosyntheseA-B.bmp : représentation schématique des deux dernières étapes de la synthèse des marqueurs membranaires des groupes sanguins ;

- phenotypeGS.bmp : texte présentant le phénotype groupes sanguins à différentes échelles.

Les documents présentent le phénotype groupes sanguins et permettent de le définir à différentes échelles :

- phénotype clinique mis en évidence par les accidents de transfusion ;
- phénotype cellulaire : marqueurs membranaires ;
- phénotype moléculaire : enzyme codée par le gène ABO (enzyme sous la forme A, B ou inactive O).

Remarques :

- il est important pour la suite de la démarche que l'élève ait bien compris que le groupe O est lié à l'absence de marqueurs A et B ;
- il est important que l'élève comprenne également que le phénotype moléculaire ne peut être le marqueur membranaire (qui n'est pas une protéine) mais l'enzyme permettant sa synthèse.

La comparaison des séquences des différentes protéines ABO et la mise en relation avec les phénotypes cliniques correspondants permettent de discuter des relations entre phénotypes moléculaire, cellulaire et clinique. Les connaissances sur les protéines enzymatiques seront mobilisées pour expliquer les observations. On montrera ainsi que les phénotypes alternatifs sont dus à des différences dans les protéines concernées.

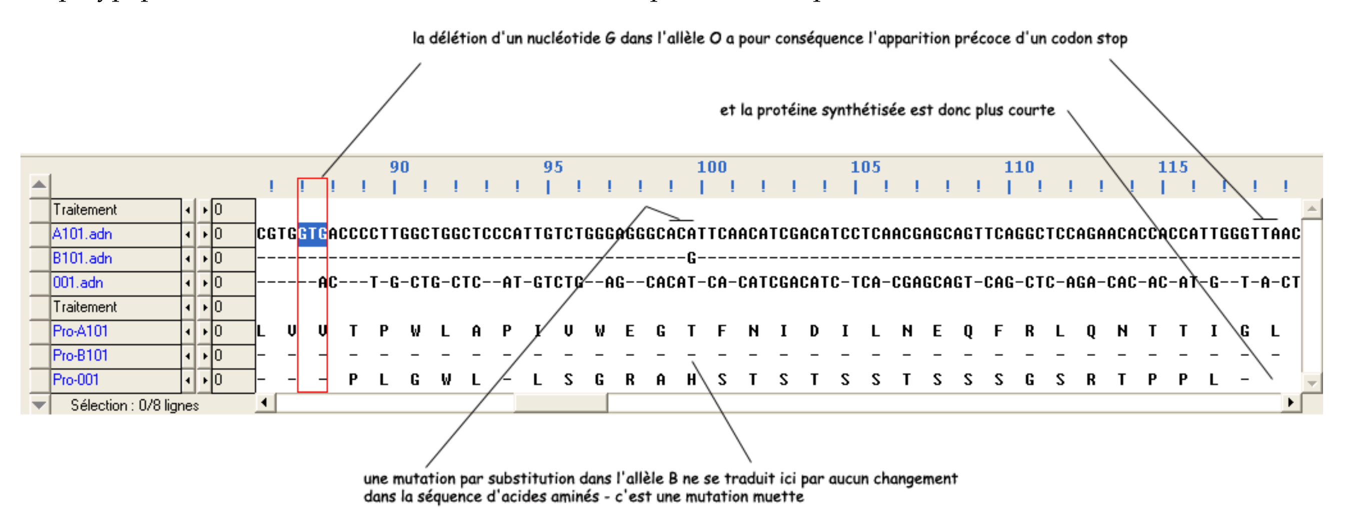
Bilan

Les enzymes EA et EB diffèrent par quatre acides aminés, en position 174, 233, 264 et 266. Ces différences sont à mettre en relation avec les différences d'activité de ces enzymes (elles ne catalysent pas la même réaction) : certains des acides aminés différents doivent se situer au niveau du site actif de l'enzyme.

L'enzyme O est beaucoup plus courte que les deux autres (115 a.a. seulement). Cette molécule protéique ne doit donc pas pouvoir prendre une forme spatiale correcte, ce qui explique son inactivité.

La mise en relation des différences observées au niveau des quelques allèles sélectionnés ici avec les différences observées au niveau protéique permet de discuter de l'effet des mutations.

Remarque pratique : la mise en relation des différences observées entre les allèles avec celles observées entre les polypeptides est facilitée si l'on demande des comparaisons simples dans les deux cas.



Bilan

L'allèle O diffère de l'allèle A par une délétion d'un nucléotide en position 261, ce qui entraîne un décalage du cadre de lecture et donc une modification importante de la séquence d'acides aminés. Cela entraîne aussi l'apparition précoce d'un codon stop, donc une chaîne polypeptidique plus courte (116 acides aminés au lieu de 353).

L'allèle B diffère de l'allèle A par sept nucléotides ; il s'agit à chaque fois de mutations par substitution. Trois de ces mutations sont des mutations muettes (positions 297, 657 et 930) ; les quatre autres sont des mutations faux-sens (positions : 526, 703, 796 et 803).

Le gène ABO est en fait très polymorphe. Le fichier polyABO.edi fournit les séquences de plusieurs autres allèles de ce gène. Un travail similaire à celui qui précède peut être fait à partir de ces allèles, les séquences protéiques correspondantes étant obtenues en utilisant la fonctionnalité **Traduire** du logiciel.

La complexité des relations gènes/phénotype

La mise en évidence de la complexité des relations génotype/phénotype peut être abordée à partir de l'exemple des deux familles proposées.

Pour chaque famille, sont mises à disposition les données suivantes :

- un arbre qui n'est là que pour préciser le phénotype des membres d'une famille donnée (et non pour initier un raisonnement en fonction des lois de l'hérédité) ;
- les séquences des deux allèles du gène ABO que possèdent tous les membres de la famille ;
- les séquences des allèles ABO considérés comme référence.

Étude de la famille 1

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes groupes sanguins** permet, par le développement de l'arborescence, d'accéder à :

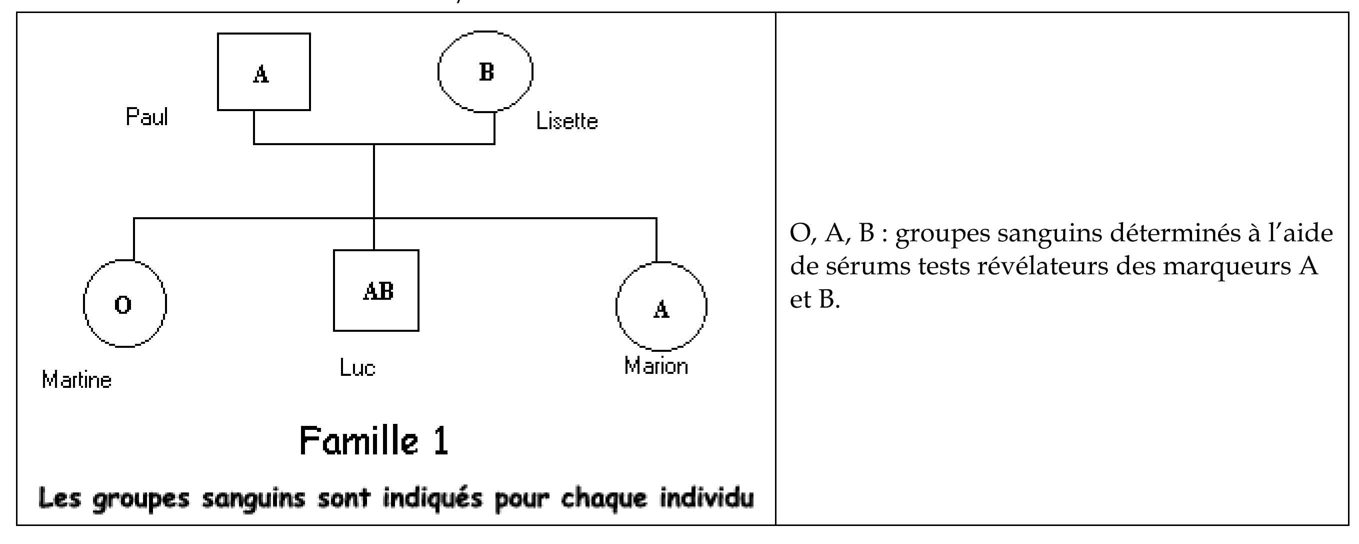
- **Nucléiques** qui charge le fichier RefABO-ADN.edi affichant les séquences nucléiques strictement codantes des allèles A101, B101 et O01 ;

- **Génotypes famille1 ABO** qui charge le fichier Fam1ABO.edi affichant les séquences nucléiques strictement codantes des allèles du gène ABO de chacun des membres de la famille 1 (génotype de Paul : A//O, de Lisette : B//O, de Martine : O//O, de Luc : A//B, de Marion : A//O).

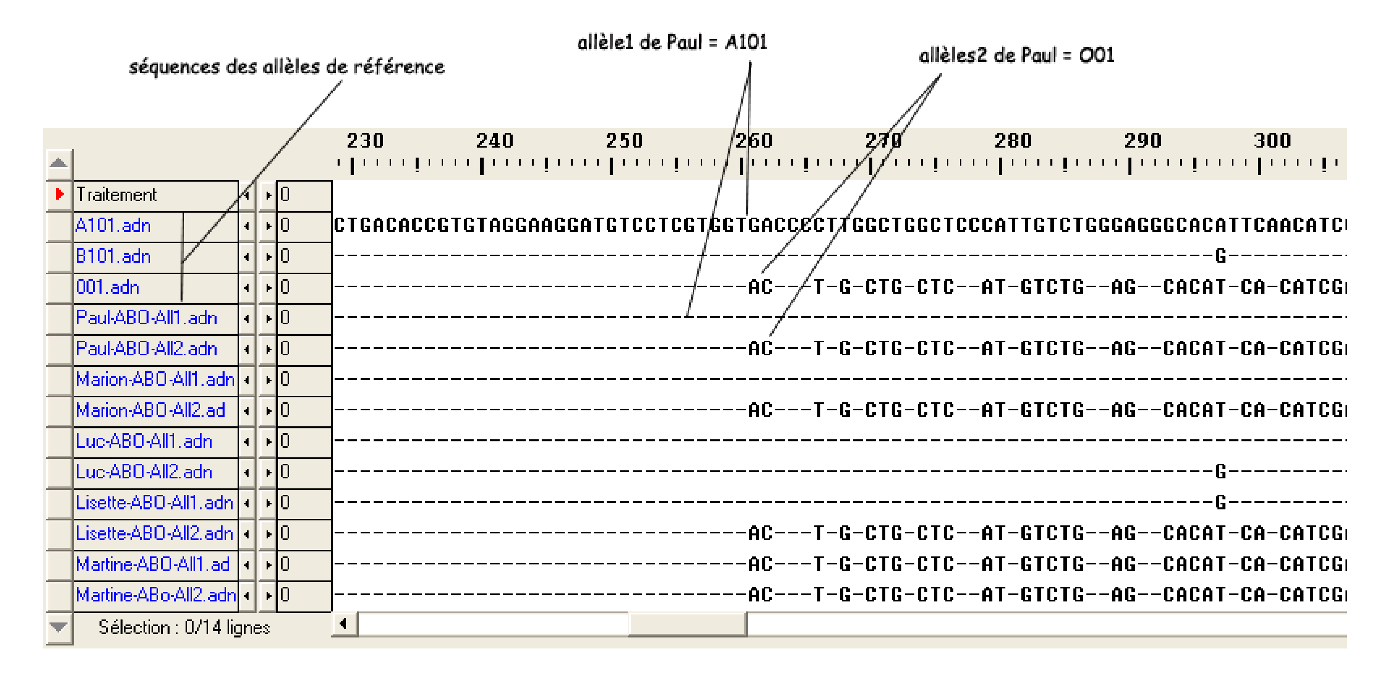
Documents fournis

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes groupes sanguins** permet d'accéder à **Dominance et récessivité (Famille 1)** qui charge le fichier famille1.bmp affichant l'arbre généalogique de la famille 1 avec indication des phénotypes groupes sanguins.

L'étude de cette famille permet de se familiariser avec la méthode de détermination des génotypes et de discuter des relations de dominance/récessivité.



La comparaison des allèles de chaque individu avec les allèles de référence du gène ABO fournis permet de déterminer le génotype de chacun des membres de la famille : le génotype de Paul est A//O, celui de Lisette est B//O, celui de Martine est O//O, celui de Luc est A//B et celui de Marion est A//O.



La mise en relation du génotype et du phénotype pour les individus hétérozygotes permet ensuite de discuter des relations de dominance/récessivité (ces relations ne peuvent se définir qu'à l'échelle du phénotype clinique, et non à l'échelle moléculaire ; en effet, les deux allèles s'expriment dans la cellule).

Étude de la famille 2

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes groupes sanguins** permet, par le développement de l'arborescence, d'accéder à :

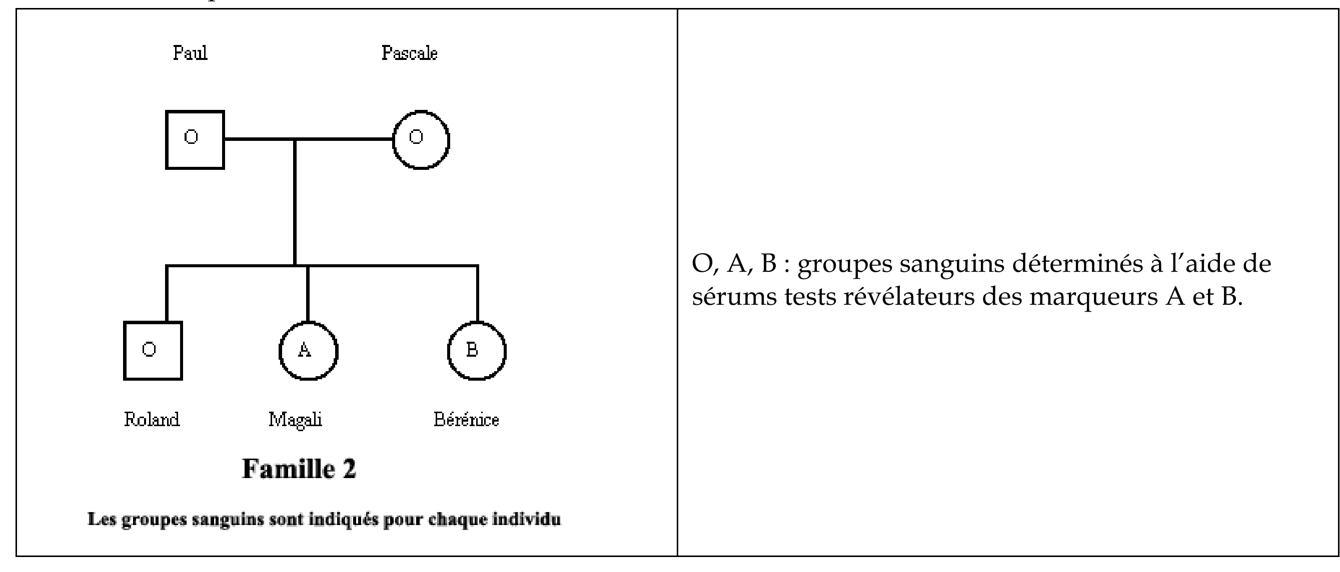
- **Nucléiques** qui charge le fichier RefABO-ADN.edi affichant les séquences nucléiques strictement codantes des allèles A101, B101 et O01 ;
- **Allèles FUT1** qui charge le fichier RefFut1.edi affichant les séquences strictement codantes des allèles de référence du gène Fut 1 (l'allèle Fut1H code pour une enzyme H fonctionnelle ; l'allèle Fut1h1 code pour une enzyme h non fonctionnelle) ;
- **Poly FUT1** qui charge le fichier polyFut1.edi affichant les séquences nucléiques strictement codantes de six allèles du gène Fut1 (allèles fut1H, codant pour une enzyme H fonctionnelle, allèles Fut1h1 à Fut1h5, codant pour des enzymes h non fonctionnelles) ;
- **Protéines FUT1** qui charge le fichier ProtFut1.edi affichant les séquences protéiques des enzymes H (fonctionnelle) et h1 (non fonctionnelle) ;
- **Génotypes famille2 ABO** qui charge le fichier Fam2ABO.edi affichant les séquences nucléiques strictement codantes des allèles du gène ABO de chacun des membres de la famille 2 (génotype de Paul : A//B, de Pascale : O//O, de Roland : B//O, de Magali : A//O, de Bérénice : B//O) ;
- **Génotypes famille2 FUT1** qui charge le fichier Fam2Fut1.edi affichant les séquences nucléiques strictement codantes des allèles du gène Fut1 de chacun des membres de la famille 2 (génotype de Paul : h//h, de Pascale : H//h, de Roland : h//h, de Magali : H//h, de Bérénice : H//h).

Documents fournis

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes groupes sanguins** permet d'accéder à **Plusieurs génotypes pour un même phénotype (Famille 2)** qui charge le fichier Famille2ABO.bmp affichant l'arbre généalogique de la famille 2 avec indication des phénotypes groupes sanguins.

Le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype groupes sanguins** permet de charger le fichier biosyntheseA-B.bmp affichant la chaîne de biosynthèse des marqueurs A et B.

L'étude de cette famille permet de découvrir un aspect de la complexité des relations génotype/phénotype : un phénotype peut dépendre de l'expression de plusieurs gènes. Il s'agit de mettre l'élève en situation de découverte du problème.



O, A, B : groupes sanguins déterminés à l'aide de sérums tests révélateurs des marqueurs A et B.

La comparaison des allèles du gène ABO de chaque individu avec les allèles de référence de ce gène permet de déterminer le génotype de chacun : le génotype de Paul est A//B et celui de Magali est A//O, celui de Pascale est O//O, et celui de Roland et de Bérénice est B//O.

Si l'on s'en tenait au fait que le phénotype groupes sanguins ne dépende que du seul gène ABO, et que l'allèle O est récessif par rapport à A et B, il y a une incohérence pour Paul et Roland ; possédant un allèle A ou B, ils

ne devraient pas être de groupe O. Il faut donc imaginer que le phénotype groupes sanguins ne dépend pas que du gène ABO.

Des renseignements complémentaires sont alors fournis concernant la synthèse des marqueurs A et B : la synthèse des marqueurs A et B ne résulte pas d'une seule réaction, mais d'une chaîne de biosynthèse. Cette chaîne nécessite plusieurs réactions chimiques catalysées chacune par une enzyme différente, codée chacune par un gène différent. Il peut donc y avoir plusieurs origines au groupe sanguin O : inactivité de l'enzyme O, mais aussi inactivité de l'enzyme H (codée par le gène Fut1), donc non-synthèse du marqueur H.

Remarque

Un travail peut être effectué sur les allèles du gène Fut1 et les protéines correspondantes pour discuter, comme il a été fait pour le gène ABO, des relations entre génotype/phénotype moléculaire/phénotype clinique. Les mutations présentes dans les allèles du gène Fut1 entraînent des modifications de la séquence d'acides aminés de l'enzyme h, responsables de l'inactivité de l'enzyme et d'un phénotype groupe sanguin O.

Résultats de comparaisons simples entre les allèles du gène Fut1 d'une part, et entre les enzymes H correspondantes d'autre part :

- Allèle Fut1h1 : diffère de l'allèle Fut1cod (allèle normal, de référence) par une substitution en position 695 – le remplacement d'un nucléotide G par un A transforme le codon TGG en codon stop TAG. Le polypeptide synthétisé est donc plus court (230 acides aminés au lieu de 399) et donc non fonctionnel ;
- Allèle Fut1h2 : diffère de l'allèle Fut1cod par une délétion d'un nucléotide en position 990 ; il s'ensuit un décalage du cadre de lecture et l'apparition précoce d'un codon stop. Le polypeptide synthétisé est donc plus court (334 acides aminés au lieu de 399) et non fonctionnel ;
- Allèle Fut1h3 : diffère de l'allèle Fut1cod par une substitution en position 721 (T → C). Il s'agit d'une mutation faux-sens qui entraîne le changement d'un acide aminé dans la séquence du polypeptide. Cette différence suffit à rendre l'enzyme non fonctionnelle ;
- Allèle Fut1h4 : diffère de l'allèle Fut1cod par une substitution en position 442 (G → T). Il s'agit d'une mutation faux-sens qui entraîne le changement d'un acide aminé dans la séquence du polypeptide. Cette différence suffit à rendre l'enzyme non fonctionnelle ;
- Allèle Fut1h5 : diffère de l'allèle Fut1cod par deux substitutions en position 460 (T → C) et 1042 (G → A). Il s'agit de mutations faux-sens qui entraînent le changement de deux acides aminés dans la séquence du polypeptide. Ces différences suffisent à rendre l'enzyme non fonctionnelle.

La comparaison des allèles du gène Fut 1 des membres de la famille 2 avec les allèles Fut1 de référence (RefFut1.edi et Fam2Fut1.edi) permet de déterminer leur génotype : Pascale, Magali et Bérénice sont de génotype H//h ; Paul et Roland sont de génotype h//h. Ainsi, l'incohérence constatée auparavant pour Paul et Roland trouve ici une explication : bien que possédant un allèle A ou B, ils sont de groupe O car la synthèse du précurseur H ne peut avoir lieu, l'enzyme h1 étant inactive.

On montre ainsi qu'un phénotype peut dépendre de plusieurs gènes, et qu'un même phénotype clinique peut correspondre à plusieurs génotypes.

Les phénotypes albinos

Informations scientifiques

Les mécanismes assurant la pigmentation de la peau

La pigmentation est un processus complexe qui résulte de la synthèse de pigments spécifiques, les mélanines, notamment au sein des mélanocytes (cellules épidermiques). Cette synthèse s'effectue dans les mélanocytes au sein d'organites appelés mélanosomes, qui migrent dans la cellule, puis sont transférés aux kératinocytes. Les kératinocytes ne font donc que stocker les mélanines et les éliminer lors de leur desquamation.

La coloration de la peau résulte donc d'une succession d'étapes : synthèse de mélanine dans les mélanosomes, transport des mélanosomes vers la périphérie des mélanocytes, passage dans les kératinocytes.

La synthèse de mélanine

Elle nécessite toute une chaîne de biosynthèse dont la matière première initiale est la tyrosine :

Biosynthèse de la mélanine

La tyrosinase est une enzyme cruciale de la chaîne de biosynthèse de la mélanine car elle catalyse quatre réactions : l'hydroxylation de la tyrosine en DOPA, l'oxydation de la DOPA en dopaquinone et les deux réactions aboutissant aux quinones.

La quantité de cette enzyme est sensiblement identique chez tous les humains, quelle que soit leur couleur de peau, mais son activité catalytique varie beaucoup d'un individu à l'autre (elle est faible chez les individus à peau claire). Cette activité dépend de l'acidité dans le mélanosome ; une diminution de l'acidité du milieu dans lequel sont cultivées des cellules de peau claire entraîne une augmentation de l'activité catalytique de la tyrosinase.

Certaines protéines sont impliquées dans la modulation de l'acidité dans les mélanosomes. Ainsi la protéine P, codée par le gène OCA2, est une protéine transmembranaire qui transporte les protons hors du mélanosome, ce qui y entraîne une baisse de l'acidité.

Le transport des mélanosomes dans les mélanocytes se fait de la zone périnucléaire où il se forme vers la périphérie de la cellule, d'où il va la quitter pour passer dans les kératinocytes. Ce déplacement est assuré par les microtubules, qui forment une sorte de tapis roulant. De nombreuses molécules sont indispensables à ce déplacement, tant au niveau des microtubules que de la membrane du mélanosome.

Le transfert des mélanosomes aux kératinocytes fait intervenir des récepteurs membranaires ; il est possible d'activer ces récepteurs dans des cultures cellulaires, et l'on obtient alors un assombrissement des kératinocytes car le transfert des mélanosomes est accru.

Les albinismes

L'albinisme humain est la plus fréquente des anomalies héréditaires de pigmentation. Le phénotype macroscopique visible (hypopigmentation de la peau, des poils, des cils, des yeux) fut pendant longtemps le seul critère retenu pour détecter cette anomalie due à une réduction (ou même à une absence) de synthèse de mélanine. Si cette hypopigmentation est générale, on parle d'albinisme oculo-cutané (OCA). Si elle est localisée aux yeux, on la nomme albinisme oculaire (OA). L'albinisme OCA est le plus courant.

Le seul critère de pigmentation est insuffisant pour diagnostiquer l'albinisme. Des changements dans les yeux et le système visuel ont été trouvés dans tous les cas d'albinisme OCA et d'albinisme OA. L'iris et la rétine étant dépigmentés, les pupilles apparaissent rouges du fait de la réflexion de la lumière sur la choroïde très riche en vaisseaux sanguins. D'autre part, le développement des voies optiques est altéré chez les individus albinos, et la plupart d'entre eux sont photophobes et présentent un nystagmus ; ils ont de plus une diminution de l'acuité stéréoscopique et un strabisme alternant. Chez une personne qui pourrait avoir une pigmentation presque normale, l'examen ophtalmologique complet permet donc de dépister un albinisme. On ne s'intéressera ici qu'à l'albinisme OCA.

On distingue deux types d'albinismes OCA, selon l'origine de l'anomalie :

- albinisme OCA1 : il met en cause une anomalie de fonctionnement de la tyrosinase (mutation dans le gène de la tyrosinase). On connaît plusieurs types d'albinisme OCA1 :
  - *l'albinisme OCA1 –1A ou « tyrosinase négative »* : ce phénotype a été trouvé dans tous les groupes ethniques à travers le monde. Il est lié à une activité nulle de la tyrosinase. Les enfants naissent avec des cheveux blancs, la peau blanche, les yeux bleus et ces caractéristiques persistent tout au long de leur vie. La peau ne bronze pas et reste très sensible au soleil tout au long de la vie ;
  - *l'albinisme OCA1 – 1B ou « type jaune »* : il est reconnu dans beaucoup de populations, et a été découvert dans la population Amish (état de l'Indiana, USA). Les enfants naissent avec les cheveux blancs et la peau blanche, mais les cheveux se colorent en blond au cours des premières années, et peuvent devenir blond foncé chez l'adulte. La peau devient crème et peut bronzer au soleil. L'iris, bleu-gris, peut aussi se pigmenter davantage ;
  - *l'albinisme OCA1 – TS ou « type thermosensible »* : nées avec une peau blanche, des poils blancs et des yeux bleus, certaines personnes, au moment de la puberté, développent une pigmentation sur les zones les plus froides de leur corps comme les bras et les jambes. Selon la température de l'endroit du corps considéré, on aura une pigmentation plus ou moins développée ;
- Albinisme OCA2 : la tyrosinase a une activité normale, et l'anomalie de pigmentation est donc due à une autre origine (mutation dans un autre gène que celui de la tyrosinase ; exemple : mutation dans le gène OCA2 codant pour la protéine P). Ce type d'albinisme est le plus courant et le plus variable. Il est moins sévère que l'albinisme OCA1, et l'hypopigmentation cutanée peut varier selon l'origine ethnique et l'hétérozygotie de l'individu. Les individus atteints de cet albinisme ont les cheveux légèrement pigmentés à la naissance, une peau blanche, des yeux bleus, et une pigmentation de la peau qui augmente avec l'âge.

Les deux types d'albinisme peuvent être distingués par le « **test du bulbe pileux** » : ce test consiste à incuber des bulbes pileux prélevés chez un individu albinos dans une solution de tyrosine, pendant vingt-quatre heures à 37 °C. Si la pigmentation redevient normale (synthèse de mélanine), on conclut à un albinisme de type OCA1, sinon il s'agit d'un albinisme de type OCA2.

Les allèles du gène de la tyrosinase

La tyrosinase résulte de l'expression d'un gène situé sur le chromosome 11.

Le séquençage du gène a montré que ce gène est très polymorphe.

Le tableau ci-dessous présente **quelques-unes des mutations à l'origine de ces allèles** :

Séquence nucléique			Polypeptide	
Allèle	Nature et position des nucléotides qui varient	Nature et position des codons qui varient	Nature et position des acides aminés changés	Type de mutation
Tyrcod1	Allèle de référence			
Tyrcod2	A → C en 575	TAT → TCT en 192	Tyr → Ser en 192	Substitution faux-sens neutre
TyralbA1	G → A en 1147	GAT → AAT en 383	Asp → Asn en 383	Substitution faux-sens
TyralbA2	A → C en 575	TAT → TCT en 192	Tyr → Ser en 192	Substitution faux-sens neutre
	Délétion de deux nucleotides TG en 732-733	TGT → TGA en 244	La protéine n'a que 243 acides aminés	Délétion
TyralbA3	A → C en 575	TAT → TCT en 192	Tyr → Ser en 192	Substitution faux-sens neutre
	G → A en 533	TGG → TAG en 178	La protéine n'a que 177 acides aminés	Substitution non-sens
TyralbA4	C → T en 242	CCT → CTT en 81	Pro → Leu en 81	Substitution faux-sens
TyralbA5	C → T en 242 A → C en 575	CCT → CTT en 81 TAT → TCT en 192	Pro → Leu en 81 Y → Ser en 192	Substitutions faux-sens
TyralbB1	A → C en 575	TAT → TCT en 192	Tyr → Ser en 192	Substitution faux-sens neutre
	C → T en 1217	CCT → CTT en 406	Pro → Leu en 406	Substitution faux-sens
TyralbTS	G → A en 1265	CGG → CAG en 422	Arg → Gln en 422	Substitution faux-sens

Caractéristiques des enzymes tyrosinase codées par ces allèles

Les allèles tyralbA1, tyralbA2, tyralbA3, tyralbA4 et tyralbA5 codent pour des enzymes totalement inactives et sont donc à l'origine d'un albinisme de type OCA-1A.

L'allèle tyralbB1 code pour une enzyme présentant une activité résiduelle, et la dopaquinone formée, qui possède une forte affinité pour les groupements sulfhydril (présents dans la cystéine) est transformée en phéomélanine (pigment jaunâtre). La voie métabolique vers la phéomélanine est donc privilégiée aux dépens de celle de l'eumélanine. Cet allèle est donc à l'origine d'un albinisme de type OCA-1B.

L'allèle tyralbTS code pour une tyrosinase inactive au-dessus de 35 °C (cette mutation serait semblable à celle qui existe chez le Chat siamois). Cet allèle est donc à l'origine d'un albinisme de type OCA1-TS.

Pistes d'exploitation pédagogiques des données fournies

L'analyse des données fournies concernant les phénotypes « albinos » permet d'aborder les notions suivantes :

- le phénotype peut se définir à différents niveaux, et le phénotype moléculaire détermine le phénotype cellulaire et macroscopique ;
- le phénotype moléculaire s'explique par le génotype ;
- un même phénotype peut correspondre à plusieurs génotypes ;
- plusieurs gènes peuvent être impliqués dans la réalisation d'un phénotype.

L'accent est particulièrement mis ici sur les deux derniers points qui illustrent la complexité des relations génotype/phénotype.



La diversité des phénotypes et les relations génotype/phénotype (chez les homozygotes)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes albinos** permet, par le développement de l'arborescence, d'accéder à :

- **Allèles tyrosinase** qui charge le fichier Alleles-tyrosinase.edi affichant les séquences strictement codantes de quelques allèles du gène de la tyrosinase (allèles tyrcod1 et tyrcod2 codant pour une enzyme à activité normale ; tyralb A1 à tyralb A5 codant pour une enzyme non fonctionnelle ; tyralb B1 codant pour une enzyme à activité réduite ; tyralb TS codant pour une enzyme thermosensible) ;
- **Enzymes tyrosinase** qui charge le fichier Protyr.edi affichant les séquences protéiques des tyrosinases correspondant aux allèles fournis.

Documents fournis

Dans la banque de documents, le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype albinos** qui permet d'accéder aux fichiers :

- OCA1.bmp affichant un texte de présentation des phénotypes cliniques de l'albinisme oculo-cutané de type 1 ;
- pigmentation.bmp affichant un texte présentant l'origine de la pigmentation de la peau ;
- biosynthMelanine.bmp présentant la chaîne de biosynthèse de la mélanine (seule la voie menant à la synthèse d'eumélanine est ici proposée), avec repérage des réactions catalysées par la tyrosinase.

Les documents fournis permettent de construire un tableau des différents niveaux de phénotypes :

	Phénotype clinique	Phénotype cellulaire	Phénotype moléculaire
Individu sain	Peau, cheveux, poils et iris normalement colorés	Synthèse normale de mélanine dans les mélanocytes	Enzyme tyrosinase fonctionnelle
Individu albinos	Peau très claire, cheveux et poils blancs, iris non coloré	Très peu de synthèse de mélanine dans les mélanocytes	Enzyme tyrosinase non fonctionnelle ou à activité très réduite

Lorsque la tyrosinase est non fonctionnelle, la synthèse de mélanine n'a pas lieu dans les mélanocytes, et la peau reste donc très claire. La mise en relation de la comparaison des allèles du **gène de la tyrosinase** avec celles des protéines correspondantes permet de discuter des relations génotype/phénotype :

- certaines mutations, bien qu'entraînant une modification de la séquence d'acides aminés de la protéine, ne modifient pas son activité (tyrcod1 et tyrcod2 codent pour des protéines fonctionnelles) ;
- certaines mutations entraînant une modification de la séquence d'acides aminés de la protéine ont pour conséquence une diminution de l'activité enzymatique (tyralb B1) ou une suppression de l'activité enzymatique (tyralb A1, tyralb A4, tyralb A5) ou encore une activité dépendante de la température (tyralb S) ;
- d'autres mutations entraînent l'apparition d'un codon stop précoce, responsable de la synthèse d'une chaîne protéique plus courte et donc d'une enzyme non fonctionnelle (tyralb A2 et tyralb A3).

De la même façon, la mise en relation de la comparaison des allèles du **gène OCA2** avec celles des protéines correspondantes permet d'aboutir aux résultats suivants :

- certaines mutations n'entraînent aucun changement de la séquence d'acides aminés de la protéine à cause de la redondance du code génétique (allèle OCA2 m4, qui diffère de l'allèle normal par un nucléotide substitué en position 963 : G → A) ;
- certaines mutations rendent la protéine inactive (la protéine P n'est pas une enzyme, mais un transporteur membranaire de petites molécules comme la tyrosine) :
    - en entraînant la synthèse d'une protéine plus courte (allèle m1 – la délétion du nucléotide G en 1960 entraîne un décalage du cadre de lecture et l'apparition précoce d'un codon stop – la protéine P m1 a une longueur de 668 acides aminés au lieu de 838) ;
    - en modifiant la séquence d'acides aminés de la protéine : allèles m2 et m3 dont la mutation d'un nucléotide par substitution (respectivement : C →G en 258, et T →G en 1418) se traduit par le changement d'un acide aminé (respectivement : Ser → Arg en 86, et Ile → Ser en 473).

La complexité des relations génotype/phénotype

La mise en évidence de la complexité des relations génotype/phénotype peut être abordée à partir de l'exemple des quatre familles proposées. Pour chaque famille, sont mises à disposition les données suivantes :

- un arbre qui n'est là que pour préciser le phénotype des membres d'une famille donnée (et non pour initier un raisonnement en fonction des lois de l'hérédité) ;
- les séquences des deux allèles du gène de la tyrosinase que possèdent tous les membres de la famille ;
- les séquences des allèles de la tyrosinase présents dans la famille ;
- pour la famille 2 sont aussi fournies les séquences des deux allèles du gène OCA2 de chaque membre de la famille et les séquences des allèles OCA2 normal et OCA2 m1 présents dans cette famille.

Familles 1, 3 et 4

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes albinos** permet, par le développement de l'arborescence, d'accéder à :

- **Allèles famille1 tyrosinase** qui charge le fichier AllelesTyrFamille1.edi affichant les séquences strictement codantes des allèles du gène de la tyrosinase des membres de la famille 1, avec les allèles du gène de la tyrosinase présents dans cette famille. Les génotypes des individus sont les suivants : I1 et II2 sont (tyralb A2//tyrcod2) – I2, II 1 et III 1 sont (tyrcod2//tyrcod2) – I 3 et II 4 sont (tyrcod1//tyralb A1) – I 4 est (tyrcod2//tyrcod1) – II 3 est (tyrcod2//tyralb A1) – III 2 et III 3 sont (tyralb A1//tyralb A2) ;

- **Allèles famille3 tyrosinase** qui charge le fichier AllelesTyrFamille3.edi affichant les séquences strictement codantes des allèles du gène de la tyrosinase des membres de la famille 2, avec les allèles du gène de la tyrosinase présents dans cette famille. Les génotypes des individus sont les suivants : I 1et III 5 sont (tyrcod1//tyrcod2) – I 2, I 3 et I 4 sont (tyrcod1//tyralb A4) – II 1 est (tyrcod1//tyralb B1) – II2 est (tyrcod2//tyralb A4) – II 3 est (tyralb A4//tyralb A4) – III 1, III 2, III 3 sont (tyralbB1//tyralb A4) – III 4 est (tyrcod2//tyralb B1) ;

- **Allèles famille4 tyrosinase** qui charge le fichier allelestyrfamille4.edi affichant les séquences strictement codantes des allèles du gène de la tyrosinase des membres de la famille 2, avec les allèles du gène de la tyrosinase présents dans cette famille. Les génotypes des individus sont les suivants : I1 est (tyrcod2//tyralb TS) – I 2 est (tyrcod1//tyralb A4) – II 1 est (tyrcod1//tyrcod2) – II2, II5 et II6 sont (tyralb A4//tyralb TS) – II 3 est (tyrcod1//tyralb TS) – II 4 est (tyrcod1//tyrcod1) – II 7 est (tyrcod1//tyralb A5) – III 1 est (tyralb A4//tyralb A5).

Documents fournis

Dans la banque des documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes albinos** permet d'accéder à :

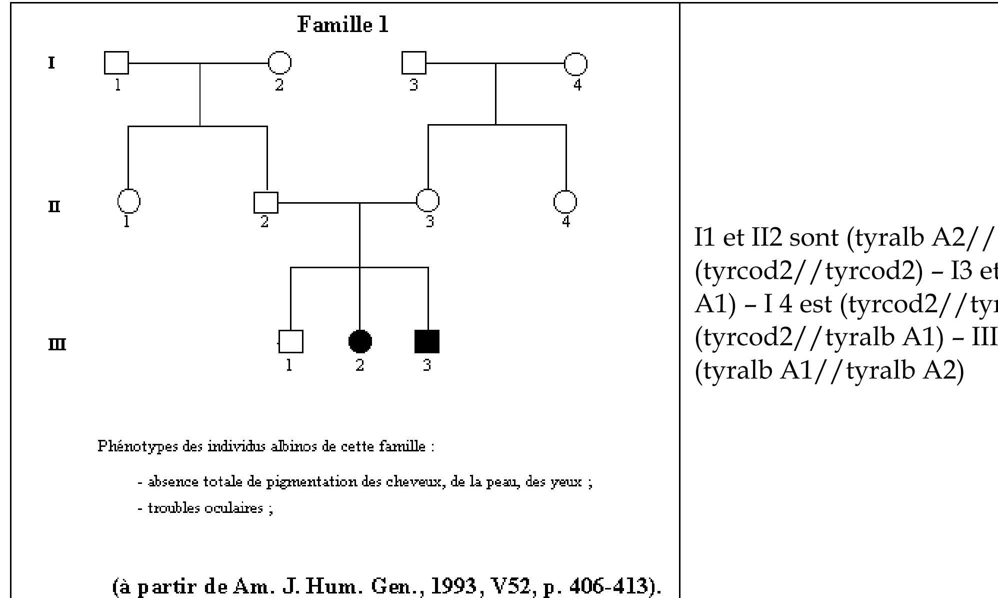
- **Polyallélisme, environnement et diversité des phénotypes (Famille 1)** pour charger le fichier arbreTyr1.bmp : arbre généalogique de la famille 1 avec indications du phénotype clinique précis des individus albinos. Dans cette famille, ce sont les allèles tyralb A1 et tyralb A2 qui sont responsables de l'albinisme ;

- **Polyallélisme, environnement et diversité des phénotypes (Famille 3)** pour charger arbreTyr3.bmp : arbre généalogique de la famille 2 avec indications du phénotype clinique précis des individus albinos. Dans cette famille, ce sont les allèles tyralb A4 et tyralb B1 qui sont responsables de l'albinisme ;

- **Polyallélisme, environnement et diversité des phénotypes (Famille 4)** pour charger arbreTyr4.bmp : arbre généalogique de la famille 4 avec indications du phénotype clinique précis des individus albinos. Dans cette famille, ce sont les allèles tyralb A4, tyralb A5 et tyralb TS qui sont responsables de l'albinisme.

Le travail peut être fait de la même façon pour chacune de ces familles, et le travail réparti entre les groupes d'élèves :

- détermination du génotype des membres de la famille, par comparaison des allèles qu'ils possèdent avec les allèles de référence fournis ;
- mise en relation génotype/phénotype, qui permet d'aboutir aux relations de dominance/récessivité entre les allèles (il y a récessivité des allèles mutés au niveau du phénotype clinique) ;
- comparaison des génotypes des individus albinos pour arriver à l'idée que plusieurs génotypes peuvent correspondre au même phénotype.

	I1 et II2 sont (tyralb A2//tyrcod2) – I2, II1 et III1 sont (tyrcod2//tyrcod2) – I3 et II4 sont (tyrcod1//tyralb A1) – I 4 est (tyrcod2//tyrcod1) – II3 est (tyrcod2//tyralb A1) – III2 et III3 sont (tyralb A1//tyralb A2)

Famille 2

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes albinos** permet, par le développement de l'arborescence, d'accéder à :

- **Allèles famille2 tyrosinase** qui charge le fichier AllelesTyrFamille2.edi affichant les séquences strictement codantes des allèles du gène de la tyrosinase des membres de la famille 2, avec les allèles du gène de la tyrosinase présents dans cette famille. Les génotypes des individus sont les suivants : I 1, III 1 et III 3 sont (tyrcod2//tyralb A4) – I 2 et II 1 sont (tyrcod2//tyralb A3) – I 3 et II 4 sont (tyrcod1//tyrcod1) – I 4, II 3 et II 5 sont (tyrcod1//tyrcod2) – II 2 est (tyralb A4//tyralb A3) – III 2 est (tyrcod1//tyralb A3) ;
- **Allèles OCA2** qui charge le fichier OCA2.edi affichant les quatre allèles du gène OCA2 et les protéines P correspondantes. L'allèle OCA2 normal code pour une protéine P fonctionnelle ; les allèles OCA2 m1, OCA2 m2 et OCA2 m3 codent pour une protéine P non fonctionnelle ;
- **Allèles famille2 OCA2** qui charge le fichier AllelesOCA2Famille2.edi affichant les allèles du gène OCA2 que possèdent les individus de la famille 2 et les deux allèles de référence présents dans cette famille (OCA2 normal et OCA2 m1). Les génotypes des individus sont les suivants : I 1, I 2, II 1, II 2 et II 5 sont (OCA2 norm//OCA2 norm) – I 3, I 4, II 3, II 4, III 1, III 2, III 3 sont (OCA2 norm//OCA2 m1) – II 3 est (OCA2 m1//OCA2 m1).

Documents fournis

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes albinos** permet d'accéder à **Plusieurs génotypes pour un même phénotype (Famille 2)** pour charger les fichiers :

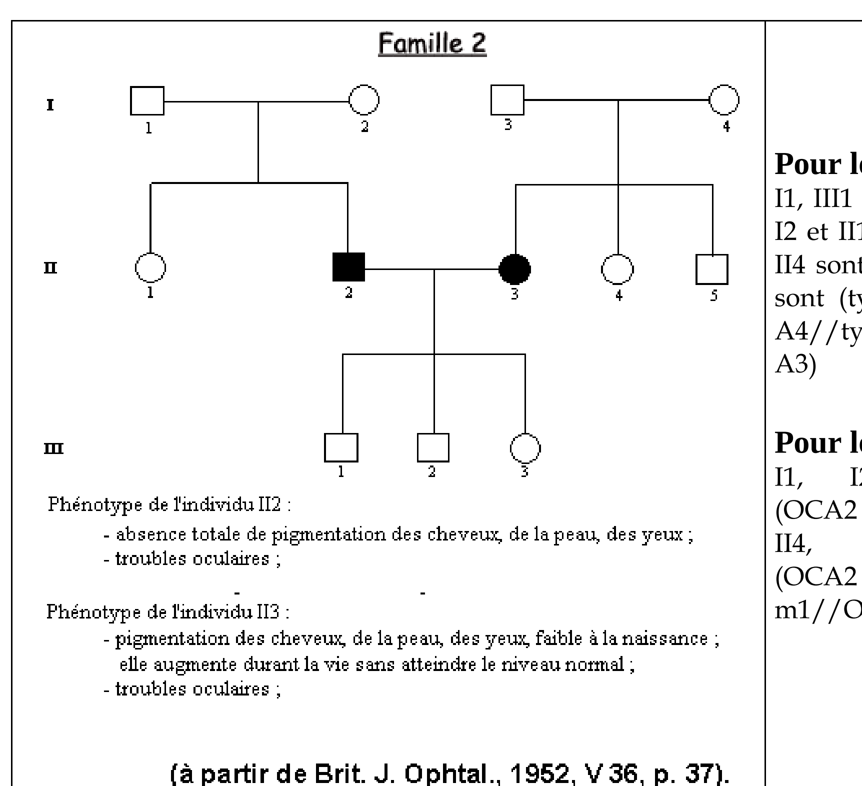
- arbreTyr2.bmp qui présente l'arbre généalogique de la famille 2 avec indications du phénotype clinique précis des individus albinos. Dans cette famille, ce sont les allèles tyralb A3 et tyralb A4 qui sont responsables de l'albinisme ;
- OCA1_2.bmp qui affiche un texte de présentation du phénotype clinique des albinismes oculo-cutané de type 1 et 2 ; de plus est présentée la description du test du bulbe pileux ;
- proteine-P.bmp qui affiche un texte introduisant le rôle de la protéine P (impliquée dans l'albinisme OCA2) ;
- pigmentation.bmp qui affiche un texte décrivant l'origine de la pigmentation de la peau.

L'étude de la famille 2 permet de mettre en évidence le fait que plusieurs gènes peuvent être impliqués dans la réalisation d'un phénotype. Les étapes de la démarche suivie peuvent être les suivantes :

- détermination du génotype des membres de cette famille pour le gène de la tyrosinase ;
- mise en relation génotype/phénotype qui doit aboutir à une incohérence pour l'individu II3 si l'on tient compte du fait que tous les allèles mutés du gène de la tyrosinase sont récessifs sur l'allèle normal. En effet, cet individu est de génotype (tyrcod1//tyrcod2) et est malade ;
- formulation d'hypothèses pour expliquer le cas de l'individu II3 (implication d'un autre gène) ;
- prise en compte d'informations nouvelles : des renseignements sur le rôle de la protéine P, qui amènent à préciser les hypothèses précédentes (l'individu II3 serait porteur de mutations dans le gène OCA2) ;
- détermination du génotype de l'individu II3 (et des autres membres de la famille) pour le gène OCA2.

Conclusion

L'individu II3 est albinos car il est porteur d'une mutation du gène OCA2 ; il ne peut donc produire une protéine P fonctionnelle, ce qui explique que la tyrosinase, bien que parfaitement normale, ne puisse être active et qu'il n'y ait donc pas de pigmentation de la peau.



Pour le gène de la tyrosinase :

I1, III1 et III3 sont (tyrcod2//tyralb A4) – I2 et II1 sont (tyrcod2//tyralb A3) – I3 et II4 sont (tyrcod1//tyrcod1) – I4, II3 et II5 sont (tyrcod1//tyrcod2) – II2 est (tyralb A4//tyralb A3) – III2 est (tyrcod1//tyralb A3)

Pour le gène OCA2 :

I1, I2, II1, II2 et II5 sont (OCA2 norm//OCA2 norm) – I3, I4, II3, II4, III1, III2, III3 sont (OCA2 norm//OCA2 m1) – II3 est (OCA2 m1//OCA2 m1)

Les phénotypes de l'alpha-antitrypsine

Informations scientifiques

Phénotype moléculaire et phénotype clinique

L'alpha AT est une protéine plasmatique constituée d'une chaîne protéique de 394 acides aminés et de trois chaînes latérales glucidiques. Ces chaînes latérales se branchent sur la chaîne peptidique au niveau de trois résidus asparagine (asn 46, 83 et 247). C'est une glycoprotéine globulaire d'une masse de 52 kdaltons. Si l'on réalise une électrophorèse des protéines plasmatiques, elle fait partie du pic des alpha1 globulines, d'où le début de son nom.

La concentration plasmatique d'alpha AT est généralement comprise entre 1,5 et 3,5 g/l. Elle diffuse dans le liquide interstitiel où sa concentration est toutefois beaucoup plus faible (environ 1/10e de celle du plasma).

La demi-vie de l'alpha AT est de quatre à cinq jours.

L'alpha AT est produite et sécrétée par les cellules hépatiques. Elle est synthétisée au niveau du REG sous forme d'un précurseur de 418 acides aminés ; le peptide signal, de 24 acides aminés, est éliminé durant le passage dans les cavités du reticulum ; c'est aussi lors de ce passage que la protéine prend sa conformation spatiale et que les trois chaînes glucidiques sont ajoutées. La protéine ainsi glycosylée est transférée dans les saccules golgiens, puis sécrétée par exocytose. On estime à 34 mg par kilogramme de masse corporelle la production journalière d'alpha AT.

L'alpha AT est un inhibiteur des protéases à sérine, protéases dont le site actif comprend la triade d'acides aminés catalytiques : acide aspartique, histidine, sérine. La trypsine, la chymotrypsine et l'élastase produite par les granulocytes sont des protéases de ce type. L'action inhibitrice de l'alpha AT sur ces protéases à sérine a été mise en évidence pour la première fois sur la trypsine, d'où son nom. *In vivo*, le seul substrat réel pour l'alpha AT est l'élastase, une endopeptidase puissante capable de cliver la plupart des protéines de la matrice extracellulaire, notamment l'élastine et les divers collagènes. Cette élastase est libérée par les granulocytes à leur mort. Au niveau du conjonctif des alvéoles pulmonaires, cette élastase est libérée en permanence à des taux très bas.

L'alpha AT protège donc la matrice extracellulaire de divers organes, et notamment au niveau pulmonaire.

Les études épidémiologiques montrent que les concentrations plasmatiques d'alpha AT inférieures à 0,8 g/l sont associées à un risque d'emphysème pulmonaire. En effet, à ces concentrations, l'inhibition de l'élastase par l'alpha AT est insuffisante. L'élastase détruit donc peu à peu le tissu conjonctif, en particulier au niveau des alvéoles pulmonaires, ce qui perturbe les échanges gazeux et entraîne l'emphysème.

L'alpha AT exerce son action protectrice en se liant fortement, de façon quasi irréversible, au site actif de l'élastase. Le site de liaison de l'alpha AT est localisé au niveau des résidus Met382 et Ser383. La méthionine peut être facilement oxydée, ce qui réduit fortement l'affinité de l'alpha AT pour l'élastase. Il semble que la fumée de cigarette entraîne l'oxydation de la méthionine, ce qui expliquerait, au moins en partie, l'aggravation des symptômes chez le fumeur en cas de déficience en alpha AT.

Le déterminisme génétique de l'alpha AT

L'alpha AT est codée par un gène situé sur le chromosome 14, et pour lequel on connaît de très nombreux allèles dans l'espèce humaine (75 allèles dont plusieurs avec une fréquence supérieure à 1 %). Ce gène est composé de sept exons et six introns et comprend 12 200 paires de bases ; la région strictement codante se trouve au niveau des quatre derniers exons. Les séquences fournies ici sont des séquences nucléiques strictement codantes, incluant la région correspondante au peptide signal (24 acides aminés), qui sera excisée dans la protéine mature.

Les allèles peuvent être regroupés en trois catégories :

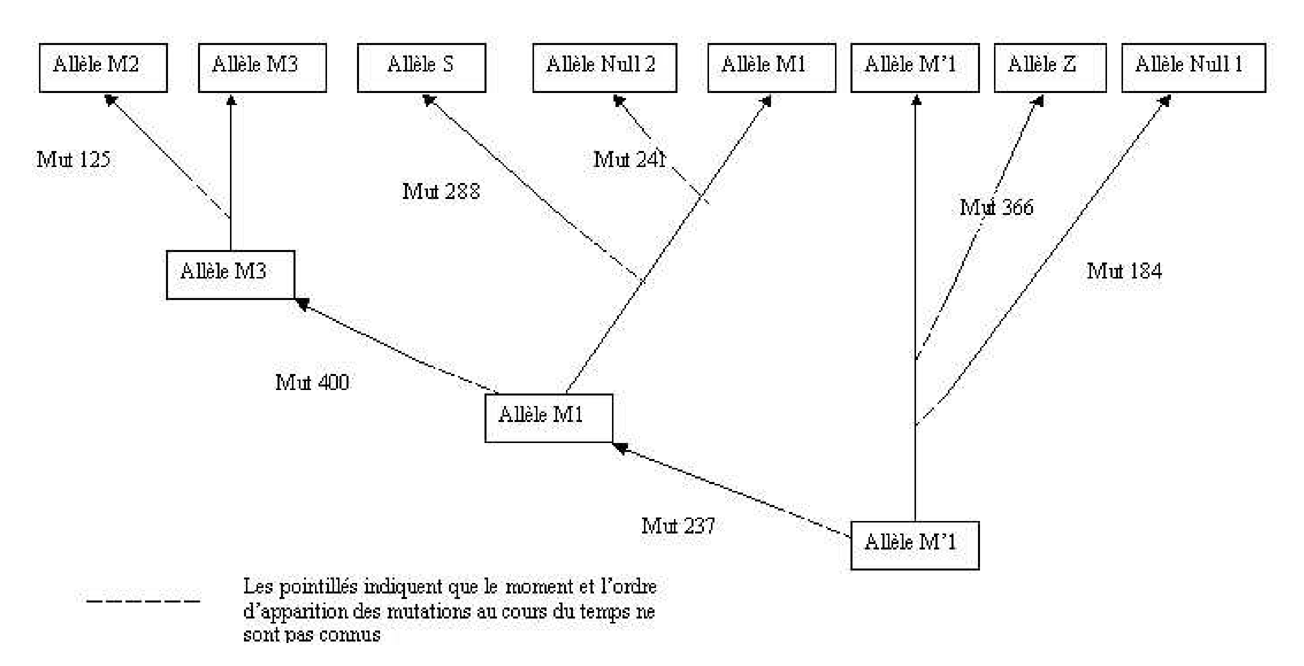
- les « variants normaux » (M'1, M1, M2, M3) : ils codent pour des molécules d'alpha AT fonctionnelles. Les mutations apparues dans les allèles M1, M2 et M3 entraînent donc des modifications de séquences d'acides aminés sans conséquences pour les propriétés de la molécule d'alpha AT (mutations neutres). Il existe de nombreux autres allèles « variants normaux », mais leur fréquence est plus faible (l'allèle M4 ayant toutefois une fréquence de 1 à 5 % dans les populations européennes) ;

- les « variants déficients » (S et Z) :
  - L'allèle S code pour une molécule d'alpha AT fonctionnelle, mais sécrétée en plus faible quantité que les molécules codées par les variants normaux. La mutation (codon 288) entraîne un changement d'acide aminé qui semble entraîner une relative instabilité de la molécule d'alpha AT, cause d'une destruction précoce à l'intérieur même des cellules hépatiques. Ce phénomène serait à l'origine de la déficience de sécrétion, la molécule d'alpha AT ayant une durée de vie normale une fois sécrétée ;
  - L'allèle Z code pour une molécule d'alpha AT ayant une activité réduite, et sécrétée en plus faible quantité. Chez les homozygotes Z//Z, les niveaux d'ARNm sont normaux, ainsi que la traduction, mais après leur mise en forme dans le REG, les molécules d'alpha AT s'agrègent ce qui limite leur transfert vers l'appareil de Golgi et donc leur sécrétion. La structure tridimensionnelle de l'alpha AT est affectée par cette substitution ;
- les « variants Null » : ces allèles sont toujours rares, avec une fréquence inférieure à 0,1 %. Les mutations apparues dans ces allèles font apparaître un codon stop précoce. Les protéines codées par ces allèles sont raccourcies, très instables et rapidement détruites.

L'évolution du gène de l'apha AT

Bien que l'allèle M1 soit le plus fréquent aujourd'hui, les généticiens estiment que l'allèle M'1 est l'allèle présent dans les premières populations humaines. Ils s'appuient pour cela sur la séquence codante du gène connue chez le chimpanzé : cette séquence est identique à celle de M'1, sauf à un site.

Si M'1 est l'allèle initial, tous les autres en dérivent par mutations. En admettant que si deux allèles possèdent la même mutation, ils ont une parenté plus grande, on peut établir une filiation entre les allèles de ce gène.



Filiation des allèles de l'alpha AT

Pistes d'exploitation pédagogique des données fournies

Les relations génotype/phénotype/environnement

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes de l'alpha-antitrypsine** permet, par le développement de l'arborescence, d'accéder à :

- **Allèles du gène AT** qui charge le fichier allAT.edi affichant les séquences nucléiques strictement codantes de quelques allèles du gène de l'alpha AT – allèles M'1, M1, M2, M3, Z, S, NULL 1 et NULL 2 ;
- **Génotypes famille1** qui charge le fichier allATfamille1.edi affichant les séquences des allèles du gène de l'alpha-antitrypsine des membres d'une famille dans laquelle existent des cas d'emphysème pulmonaire. Les allèles présents dans cette famille sont M'1, M1, NULL 1 et Z. Un frère et une sœur de cette famille ont la même concentration plasmatique en alpha AT, mais seul le frère est atteint d'emphysème pulmonaire ; il est précisé sur l'arbre que cet Homme est fumeur.

Documents fournis

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Les phénotypes de l'alpha-antitrypsine** permet d'accéder aux fichiers :

- InfoAlphaAT.bmp pour afficher des informations scientifiques, sous forme d'un petit texte, concernant l'alpha ATet la relation entre le déficit en alpha AT fonctionnelle et l'emphysème pulmonaire ;
- familleAlphaAT.bmp qui présente l'arbre généalogique d'une famille présentant des cas d'emphysème pulmonaire. Les phénotypes cliniques et biochimiques (concentrations plasmatiques en alpha AT) sont indiqués pour chaque individu. Les individus fumeurs sont indiqués ;
- AllellesAlphaAT.bmp qui présente, sous la forme d'un tableau, la fréquence estimée de chaque allèle dans la population blanche des États-unis d'Amérique, et pour chacun de ces allèles indique la concentration d'alpha AT dans le sang et le risque de maladie chez l'homozygote ;
- GenoPhenoAlphaAT.bmp qui affiche un tableau indiquant les phénotypes biochimiques (concentration en alpha AT) et cliniques en fonction des différents génotypes ;
- repartitionSZ.bmp qui affiche les cartes de distribution de fréquence de l'allèle S et de l'allèle Z en Europe.

Résultats obtenus par l'exploitation des séquences nucléiques fournies

Allèle	Différence par rapport à l'allèle de référence (M'1)		Différence entre la protéine codée par M'1 et la protéine codée par cet allèle
	Nucléotide	Codon	
M1	710 : C → T	Codon 237 : GCG → GTG	A → V
M2	374 : G → A 710 : C → T 1200 : A → C	Codon 125 : CGT → CAT Codon 237 : GCG → GTG Codon 400 : GAA → GAC	R → H A → V E → D
M3	710 : C → T 1200 : A → C	Codon 237 : GCG → GTG Codon 400 : GAA → GAC	A → V E → D
S	710 : C → T 863 : A → T	Codon 237 : GCG → GTG Codon 288 : GAA → GTA	A → V E → V
Z	1096 : G → A	Codon 366 : GAG → AAG	E → K
Null 1	552 : délétion de C	Codon 184 : TAC → TAG	183 acides aminés au lieu de 418
Null 2	710 : C → T 721 : A → T	Codon 237 : GCG → GTG Codon 241 : AAG → TAG	A → V 240 acides aminés au lieu de 418

Tableau de comparaison des allèles de l'alpha AT et des protéines correspondantes

La comparaison des allèles des membres de cette famille aux différents allèles du gène de l'alpha AT permet de déterminer le génotype de chaque individu de la famille. L'arbre généalogique fourni précisant les taux

d'alpha AT plasmatique des différents membres de la famille ainsi que leur phénotype clinique, on pourra ainsi discuter de la relation génotype/phénotype.

Arbre généalogique d'une famille dont certains membres présentent des cas d'emphysème pulmonaire

I — 1 : 1,3 g/l ; 2 : 1,15 g/l

II — 1 : 0,3 g/l ; 2 : 1,10 g/l ; 3 : 1,65 g/l ; 4 : 2,35 g/l ; 5 : 0 g/l ; 6 : 2,9 g/l

III — 1 : 1,4 g/l ; 2 : 1,02 g/l ; 3 : 0,9 g/l

: emphysème pulmonaire

: pas de signes d'emphysème pulmonaire

Individu II5 : décédé à 40 ans. Troubles depuis l'âge de 30 ans

Individu III3 : fumeur.

Génotypes des individus de la famille

Individu	Génotype
I 1	M'1// Null 1
I 2	M1// Null 1
II 1	Z// Z
II 2	M'1// Null 1
II 3	M'1// Null 1
II 4	M'1// M1
II 5	Null 1// Null 1
II 6	M1// M1
III 1	M'1// Z
III 2	M1// Null 1
III 3	M1// Null 1

Les deux allèles présents chez un individu s'expriment indépendamment l'un de l'autre. On trouve donc dans le sang, en cas d'hétérozygotie, les deux types d'alpha AT. On peut donc parler de codominance à ce niveau. Cependant, au niveau clinique, on peut parler de phénotype normal dominant sur les phénotypes S, Z, Null 1 et Null 2. Ainsi :

- les personnes qui possèdent deux allèles Null présentent les symptômes d'emphysème avant l'âge de 30 ans et vivent rarement au-delà de 40 ans ;
- les individus de génotype Z//Z, donc qui possèdent des concentrations plasmatiques variant entre 0,15 et 0,5 g/l d'alpha AT souffrent aussi d'emphysème, mais plus tardivement et leur espérance de vie peut atteindre 60 ans ;
- pour les individus ayant une concentration en alpha AT à la limite du risque d'emphysème (0,9 à 1,2 g/l), le risque peut être fortement aggravé par des facteurs de l'environnement comme la fumée du tabac.

La possession d'un seul allèle normal suffit pour avoir une concentration suffisante d'alpha AT protectrice vis-à-vis de l'élastase.



Xeroderma

Informations scientifiques

Le phénotype xérodermique ou *Xeroderma pigmentosum*

Au niveau clinique : les symptômes de la maladie

Le *Xerodema pigmentosum* se traduit par des anomalies de pigmentation de la peau. Les premiers signes apparaissent très précocement, dès l'âge d'1 ou 2 ans. On observe d'abord une hypersensibilité aux rayons UV solaires (provoquant des érythèmes intenses), puis des altérations de la peau exposée (sécheresse cutanée, taches hyperpigmentées, kératose). C'est donc une affection radio-induite.

Des inflammations de la cornée et de la conjonctive apparaissent dès l'âge de 4 ans. Des tumeurs cutanées et ophtalmiques sont fréquentes dès l'âge de 8 ans, et la fréquence des cancers cutanés est multipliée par presque 5 000 chez les sujets de moins de 20 ans, par rapport à un groupe témoin. Environ 20 % des malades développent aussi des anomalies neurologiques, avec une perte progressive des neurones du cortex cérébral.

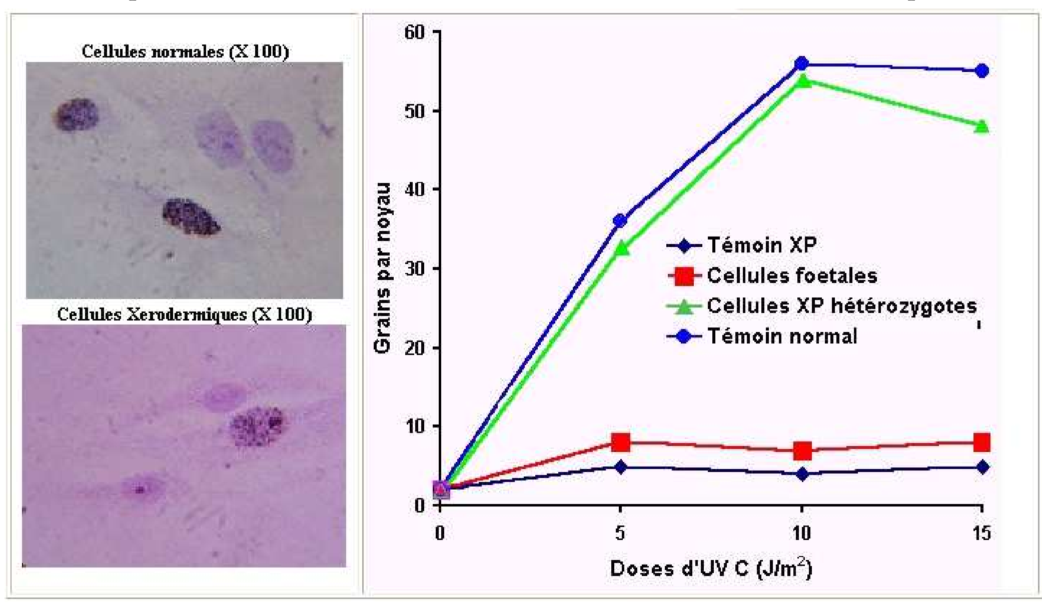
L'espérance de vie des malades est réduite de 30 ans en moyenne par rapport à un groupe témoin.

Au niveau cellulaire : les caractéristiques des cellules des malades

Les troubles observés résultent de l'atteinte de l'ADN des cellules (surtout des cellules cutanées) par les rayons UV (UVB surtout).

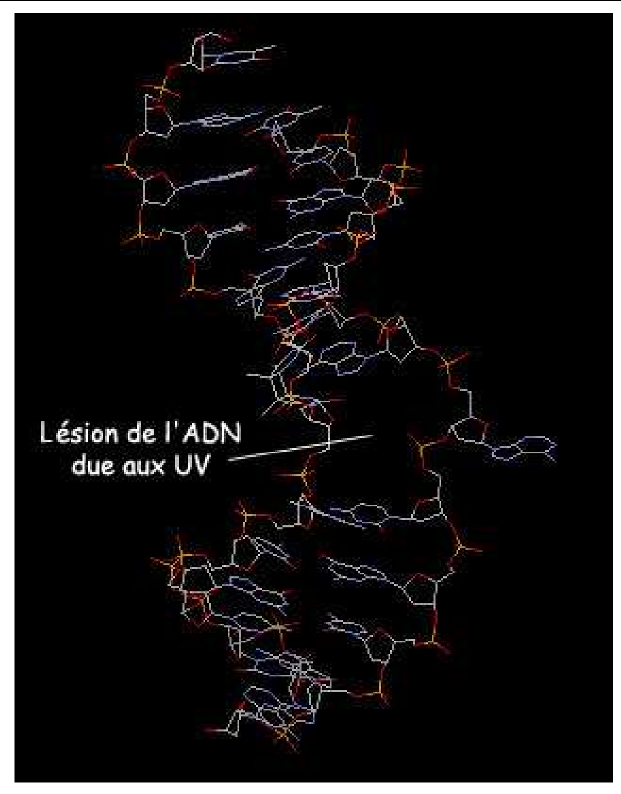
Une expérience permet de mettre en évidence les mécanismes déficients dans les cellules d'un individu xérodermique. La technique présentée ci-dessous est utilisée pour le diagnostic anténatal du *Xeroderma pigmentosum* (elle se fait alors sur des cellules prélevées par biopsie du trophoblaste à la 9e semaine ou sur des cellules prélevées lors d'une amniocentèse à la 16e semaine) :

- protocole : des cellules préalablement irradiées aux UV sont mises en présence de thymidine tritiée ; la mesure est effectuée par autoradiographie en comptant le nombre de grains d'argent par noyau. Cette méthode de diagnostic est appelée méthode UDS (*Unscheduled DNA Synthesis*) ;
- résultats : on constate une incorporation de thymidine tritiée beaucoup plus importante dans les cellules issues d'individus sains que dans les cellules des malades atteints de *Xeroderma*, comme l'indique le document ;



Le diagnostic anténatal de Xeroderma pigmentosum *: résultats par la méthode de l'UDS*

- interprétation : l'irradiation aux UV a entraîné l'apparition de nombreuses mutations dans l'ADN des cellules, notamment la formation d'une liaison covalente entre deux pyrimidines adjacentes sur un même brin ; il y a donc formation de dimères TT, CC, CT, TC. La formation de ces dimères a pour conséquence une distorsion de la double hélice (forte courbure de 7 à 44°), avec des répercussions sur la transcription, la réplication. Le document suivant présente une molécule d'ADN ainsi altérée (image obtenue à l'aide du logiciel *RasTop* – fichier ADNlese.pdb).

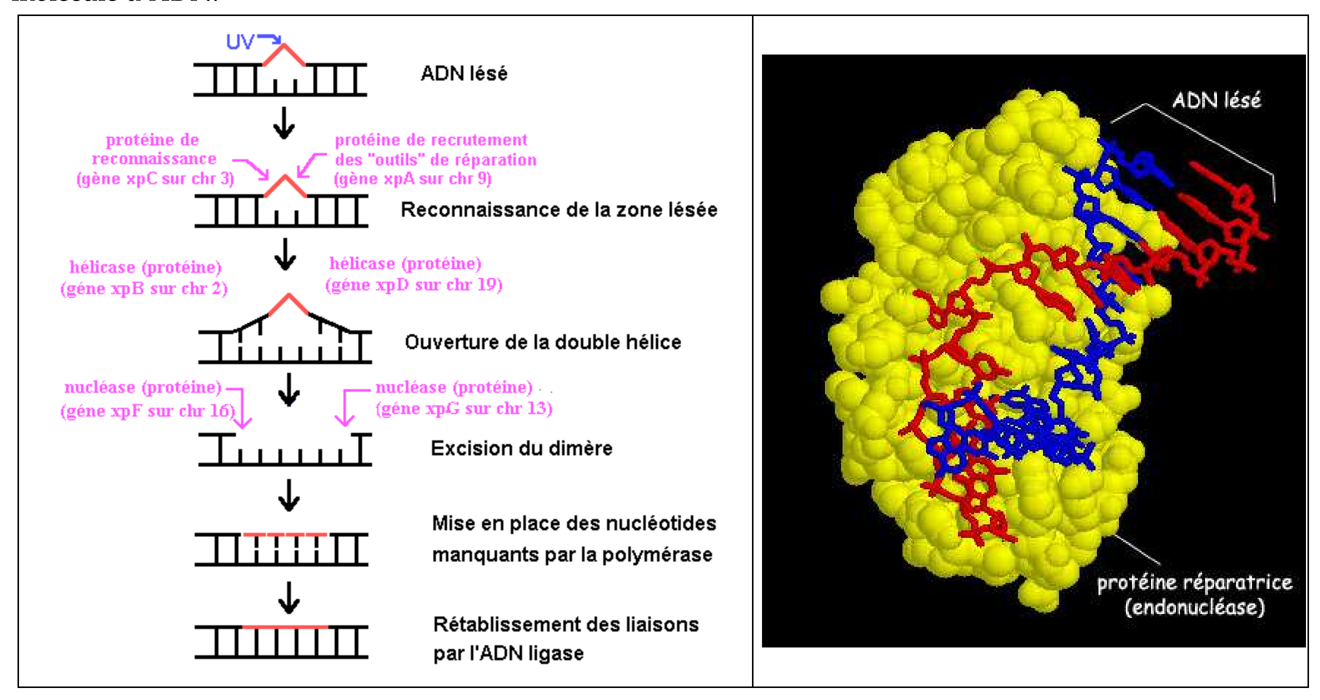


Dans les cellules normales, ces mutations sont détectées et réparées selon le mécanisme présenté plus loin ; il y a donc incorporation de thymidine tritiée pour la synthèse de l'ADN reconstruit. Dans les celllules de l'individu atteint de *Xeroderma*, cette incorportation n'a pas lieu, ce qui laisse penser que la réparation de l'ADN ne se fait pas correctement.

Habituellement, 80 % de l'ADN est réparé en quinze minutes ; chez les individus atteints de *Xeroderma,* le taux de réparation n'est que de 10 %.

Le phénotype cellulaire Xeroderma consiste donc en une hypersensibilité aux rayons UV due à un défaut dans le processus de réparation des lésions de l'ADN provoqués par les rayons UV de la lumière solaire.

Le schéma suivant présente de façon simplifiée les mécanismes réparateurs de l'ADN lésé, et l'image à droite (obtenue avec le logiciel *RasTop*, fichier ADN-Prot-1vas.pdb) présente une protéine réparatrice fixée sur la molécule d'ADN.



Déterminisme génétique du phénotype moléculaire

Chez l'Homme, une trentaine de protéines sont impliquées. Seules quelques-unes d'entre elles seront étudiées ici :

	Protéines impliquées	Gène impliqué
Recherche et reconnaissance de la lésion	• XpC permet la fixation de la protéine xpA sur l'ADN. • xpA est une protéine de 273 acides aminés à quatre doigts de zinc impliqués dans la liaison à l'ADN ; elle participe au recrutement des autres protéines	Le gène xpC est localisé sur le chromosome 3 Le gène xpA est situé sur le chromosome 9
Ouverture de la double hélice	• 3xpD (ou ERCC2) et xpB (ou ERCC3), protéines ayant respectivement 760 et 782 acides aminés. • Elles participent à l'ouverture de la double hélice sur environ 25 pb	Le gène xpD est situé sur le chromosome 19 Le gène xpB est situé sur le chromosome 2
Incision de part et d'autre de la lésion et excision	• xpG (ou ERCC5) et xpF (ou ERCC4), protéines ayant respectivement 1186 et 916 acides aminés. • Ce sont des endonucléases qui coupent respectivement en 3' et 5' de la lésion, incisant ainsi un fragment de 27 à 29 nucléotides. Elles clivent spécifiquement la molécule d'ADN au niveau où les deux brins se désolidarisent.	Le gène xpG est situé sur le chromosome 13 Le gène xpF est situé sur le chromosome 16
Resynthèse puis ligation	• Plusieurs protéines sont impliquées, et interviennent notamment dans le recrutement des ADN polymérases qui vont « combler la brèche ».	

Chez les individus atteints de *Xeroderma pigmentosum*, une des premières étapes de cette réparation est déficiente. Au moins huit gènes seraient impliqués dans cette maladie.

Relation génotype, phénotype, environnement

La seule solution thérapeutique reste la prévention : il s'agit de maintenir le malade à l'abri de toute lumière solaire. Les sujets atteints qui sortent dehors le jour doivent porter des vêtements à manches longues, des pantalons longs, un chapeau à larges bords et des lunettes de soleil ; le visage et les mains doivent être protégés par une crème spéciale, à base d'oxyde de zinc ou d'oxyde de titane ; les textiles utilisés ne doivent pas laisser passer la lumière. À l'intérieur, durant la journée, ils doivent se tenir loin des fenêtres, et la lumière artificielle doit être émise par des lampes munies de filtres.

Les tumeurs sont enlevées au fur et à mesure de leur apparition, et on utilise des composés rétinoïdes pour empêcher leur progression. Des autogreffes de peau prélevée dans les parties non exposées au soleil sont également réalisées.

Pistes d'exploitation pédagogique des données fournies

L'exploitation des données fournies permet d'aborder les notions suivantes :

- le phénotype peut se définir à plusieurs niveaux ;
- le génotype détermine le phénotype moléculaire, lui-même responsable du phénotype cellulaire et clinique ;
- plusieurs génotypes peuvent correspondre au même phénotype ;
- il existe des relations de dominance/récessivité entre les allèles d'un gène ;
- des facteurs de l'environnement peuvent avoir une influence sur le phénotype.

Des données complémentaires sont disponibles sur le site *Biotic* de l'INRP (notamment en ce qui concerne la thérapie génique). D'autre part, il est à noter que les données fournies ici peuvent aussi être utilisées dans le cadre de l'étude du polymorphisme génique.

 **La diversité des phénotypes et les relations génotype/phénotype**

Documents fournis

Dans la banque de documents, le chemin **Relations génotype-phénotype à différents niveaux d'organisation du vivant/Le phénotype xeroderma** permet d'accéder aux fichiers :

- phenoxeroderma.bmp qui affiche un texte de présentation du phénotype xérodermique et des gènes mis en cause ;

- ADNanormal.bmp qui présente une visualisation 3D obtenue avec le logiciel *RasTop* (fichier ADNlese.pdb) d'un fragment d'ADN altéré par des UV (formation d'un dimère entre deux bases pyrimidiques adjacentes, ce qui provoque une déformation de l'ADN) ;

- protreparADN.bmp qui présente une visualisation 3D obtenue avec le logiciel *RasTop* (ADN-Prot-1vas.pdb) montrant une endonucléase positionnée sur l'ADN à réparer ;

- repar.bmp qui affiche un schéma présentant de façon simplifiée les étapes d'un des mécanismes cellulaire de réparation de l'ADN.

Fichiers des molécules en 3D

Dans le dossier 3D sous Anagene2, le répertoire **Xeroderma** contient les fichiers .pdb suivants : ADNlese.pdb et ADN-Prot-1vas.pdb.

L'étude de ces documents permet de définir le phénotype à différents niveaux

Phénotype clinique	Phénotype cellulaire	Phénotype moléculaire
Individu sain	Réparation rapide et efficace des altérations de l'ADN	Protéines du système de réparation de l'ADN fonctionnelles (protéines xpa, xpb…)
Individu atteint de Xeroderma	Réparation très réduite des altérations de l'ADN qui s'accumulent donc dans les cellules	Certaines protéines du système de réparation de l'ADN ne sont pas fonctionnelles

Le génotype détermine le phénotype moléculaire, lui-même responsable du phénotype cellulaire et clinique

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Xeroderma** permet, par développement de l'arborescence, d'atteindre **Polymorphisme xpa** pour accéder à **Allèles xpa** qui charge le fichier xpa_adn.edi affichant les séquences strictement codantes de 10 allèles du gène xpa (allèles xpa_Norm et xpa_1 à xpa_9).

Les résultats et conclusions des comparaisons entre les allèles d'un gène d'une part et les protéines correspondantes d'autre part sont fournies ci-dessous.

	Différences au niveau nucléique		**Différences au niveau protéique**
	Nucléotide modifié	**Codon modifié**	
Xpa Norm	Allèle de référence		
Xpa_1	En 216 G → A	Codon 72 AGG → AGA	Aucun changement de la séquence
Xpa_2	En 149 G → T	Codon 50 TGT → TTT	C → F
Xpa_3	En 381 G → C	Codon 127 CAG → CAC	Q → H
Xpa_4	En 457 C → T	Codon 153 CGA → TGA (codon stop)	Chaîne plus courte : 152 acides aminés au lieu de 215
Xpa_5	En 557 A → G	Codon 186 CAT →CGT	H → R
Xpa_6	En 139 T → G En 148 T → A En 202 T → G En 211 T → A	Codon 47 TGC → GGC Codon 50 TGT → AGT Codon 68 TGT → GGT Codon 71 TGC → AGC	C → G C→ S C → G C → S
Xpa_7	Délétion des nucléotides 59 à 78	Décalage du cadre de lecture	Chaîne plus courte : 35 acides aminés au lieu de 215
Xpa_8	Délétion du nucléotide C 200	Décalage du cadre de lecture Apparition d'un codon stop précoce	Chaîne plus courte : 80 acides aminés au lieu de 215
Xpa_9	Délétion des nucléotides 172 à 176	Décalage du cadre de lecture	Chaîne plus courte : 61 acides aminés au lieu de 215

Les protéines xpaNorm et xpa 1 sont fonctionnelles. Les autres protéines ont une activité nulle ou réduite.

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Xeroderma** permet, par développement de l'arborescence, d'atteindre **Polymorphisme xpb** pour accéder à **Allèles xpb** qui charge le fichier xpb_adn.edi affichant les séquences strictement codantes de 3 allèles du gène xpb (allèles xpb_Norm et xpb_1 et xpb_2).

	Différences au niveau nucléique		**Différences au niveau protéique**
	Nucléotide modifié	**Codon modifié**	
Xpb Norm	Allèle de référence		
Xpb_1	En 296 T → C	Codon 99 TTC → TCC	F → S
Xpb_2	En 355 A → C	Codon 119 ACT → CCT	T → P

La protéine xpbNorm est fonctionnelle. Les autres protéines ont une activité nulle ou réduite.

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Xeroderma** permet, par développement de l'arborescence, d'atteindre **Polymorphisme xpc** pour accéder à Allèles xpc qui charge le fichier xpc_adn.edi affichant les séquences strictement codantes de quatre allèles du gène xpc (allèles xpc_Norm et xpc_1 à xpc_3).

	Différences au niveau nucléique		**Différences au niveau protéique**
	Nucléotide modifié	**Codon modifié**	
Xpc Norm	Allèle de référence		
Xpc_1	En 2452 A → C	Codon 818 AAA → CAA	K → Q
Xpc_2	Addition d'un triplet GTG en 1744, 1745, 1746	Codon 582 en plus	Un acide aminé en plus en 582 ; chaîne de 941 acides aminés au lieu de 940
Xpc_3	Délétion de deux nucléotides en 1137 et 1138	Décalage du cadre de lecture	Chaîne plus courte ; 401 acides aminés au lieu de 940

La protéine xpcNorm est fonctionnelle. Les autres protéines ont une activité nulle ou réduite.

Fichiers des séquences
Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Xeroderma** permet, par développement de l'arborescence, d'atteindre **Polymorphisme xpd** pour accéder à Allèles xpd qui charge le fichier xpd_adn.edi affichant les séquences strictement codantes de six allèles du gène xpd (allèles xpd_Norm et xpd_1 à xpd_5).

	Différences au niveau nucléique		Différences au niveau protéique
	Nucléotide modifié	**Codon modifié**	
Xpd Norm	Allèle de référence		
Xpd_1	En 1975 C → T	Codon 659 CGG → TGG	R → W
Xpd_2	En 1553 A → G	Codon 518 TAC → TGC	Y → C
Xpd_3	En 1976 G → A	Codon 659 CGG → CAG	R → Q
Xpd_4	En 2104 C → T	Codon 702 CAG → TAG (codon stop)	Chaîne plus courte : 701 acides aminés au lieu de 707
Xpd_5	En 1309 C → G	Codon 437 CTG → GTG	L → V
	Délétion à partir du nucléotide 2078	Manque tous les codons à partir du 693	Chaîne plus courte : 692 acides aminés au lieu de 707

La protéine xpdNorm est fonctionnelle. Les autres protéines ont une activité nulle ou réduite.

Fichiers des séquences
Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Xeroderma** permet, par développement de l'arborescence, d'atteindre **Polymorphisme xpf** pour accéder à à Allèles xpf qui charge le fichier xpf_adn.edi affichant les séquences strictement codantes de sept allèles du gène xpf (allèles xpf_Norm et xpf_1 à xpf_6).

	Différences au niveau nucléique		Différences au niveau protéique
	Nucléotide modifié	**Codon modifié**	
Xpf Norm	Allèle de référence		
Xpf_1	En 2362 C → T	Codon 788 CGG → TGG	R → W
Xpf_2	1471 G → A 1553 T → C	Codon 491 GAA → AAA 516 ATT → ACT	E → K T → I
Xpf_3	En 1666 G → A Délétion du nucléotide 1937	Codon 556 ACA → GCA Décalage du cadre de lecture	T → A Chaîne plus courte ; 672 acides aminés au lieu de 905
Xpf_4	En 1332, 1333 addition de deux nucléotides	Décalage du cadre de lecture	Chaîne plus courte : 481 acides aminés au lieu de 905
Xpf_5	En 642 A → G En 1504 G → A	Codon 214 ATA → AGA Codon 582 GCA → AGA	I → M G → R
Xpf_6	En 1436 G → A En 1790 T → C	Codon 479 CGG → CAG Codon 597 CTT → CCT	R → Q L → P

La protéine xpfNorm est fonctionnelle. Les autres protéines ont une activité nulle ou réduite.

Fichiers des séquences
Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Xeroderma** permet, par développement de l'arborescence, d'atteindre **Polymorphisme xpg** pour accéder à Allèles xpg qui charge le fichier le xpg_adn.edi affichant les séquences strictement codantes de trois allèles du gène xpg (allèles xpg_Norm et xpg_1 et xpg_2).

	Différences au niveau nucléique		Différences au niveau protéique
	Nucléotide modifié	**Codon modifié**	
Xpg Norm	Allèle de référence		
Xpg_1	En 2375 C → T	Codon 792 GCG → GTG	A → V
Xpg_2	En 3076 G → T	Codon 1026 GAG → TAG (codon stop)	Chaîne plus courte ; 1025 acides aminés au lieu de 1185

La protéine xpgNorm est fonctionnelle. Les autres protéines ont une activité nulle ou réduite.

Bilan
Des mutations nombreuses et variées pour chacun des gènes impliqués, donc un polymorphisme important.

Type de mutation	Allèles
Substitution muette	Xpa1
Substitution faux-sens	xpa2, xpa3, xpa5, xpa6, xpc1, xpd1, xpd2, xpd3, xpf1, xpf2, xpf5, xpf6, xpg1
Substitution non-sens	xpa4, xpd4, xpg2
Délétions	xpa7, xpa8, xpa9, xpc3, xpd5, xpf3
Addition	xpc2, xpf4

Des mutations aux effets variés sur le phénotype moléculaire :

- aucune variation de la séquence d'acides aminés (mutation muette) ;
- changement de la séquence d'acides aminés (mutations faux-sens ; additions) ;
- protéine plus courte (délétions, mutations non-sens) ;
- des conséquences sur le phénotype cellulaire et clinique :
    - une protéine plus courte sera non fonctionnelle ; donc le système de réparation des altérations de l'ADN sera défaillant et les cellules vont devenir facilement cancéreuses ;
    - une protéine dont la séquence change beaucoup sera aussi non fonctionnelle (le rôle d'une protéine dépend de sa structure spatiale, elle-même déterminée par sa séquence), avec les mêmes conséquences que précédemment ;
    - le changement d'un seul acide aminé aura des conséquences variables : elles seront importantes si cet acide aminé fait partie du site actif ou est important pour la structure spatiale de la protéine.

 La complexité des relations génotype/phénotype

Séquences et documents

Fichiers des séquences

La commande **Xeroderma** de la banque de thèmes d'étude permet, par développement de l'arborescence, d'accéder à :

- **Génotypes famille1** qui charge le fichier allfam1xero.edi affichant les séquences strictement codantes des allèles du gène xpc des individus de la famille 1, et les séquences des allèles du gène xpc présents dans cette famille. Le génotype d'Adrien, Betty et Jérôme est xpcNorm//xpc1 ; celui de Priscille est xpcNorm//xpcNorm ; celui de Nicolas est xpc1//xpc1 ;
- **Génotypes famille2** qui charge le fichier allfam2xero.edi affichant les séquences strictement codantes des allèles du gène xpc des individus de la famille 1 et les séquences des allèles du gène xpc présents dans cette famille. Le génotype de Grégoire, Marie, Jean et Anne est xpcNorm//xpc2 ; celui de Mélanie est xpc2//xpc2 ;
- **Génotypes famille3** qui charge le fichier allfam3xero.edi affichant les séquences strictement codantes des allèles du gène xpc des individus de la famille 1 et les séquences des allèles du gène xpc présents dans cette famille. Michel et Sylvie ont pour génotype : xpcNorm//xpc3 ; Agnès et Tristan ont pour génotype : xpc3//xpc3 ; Elodie a pour génotype xcpNorm//xpcNorm.

Documents fournis

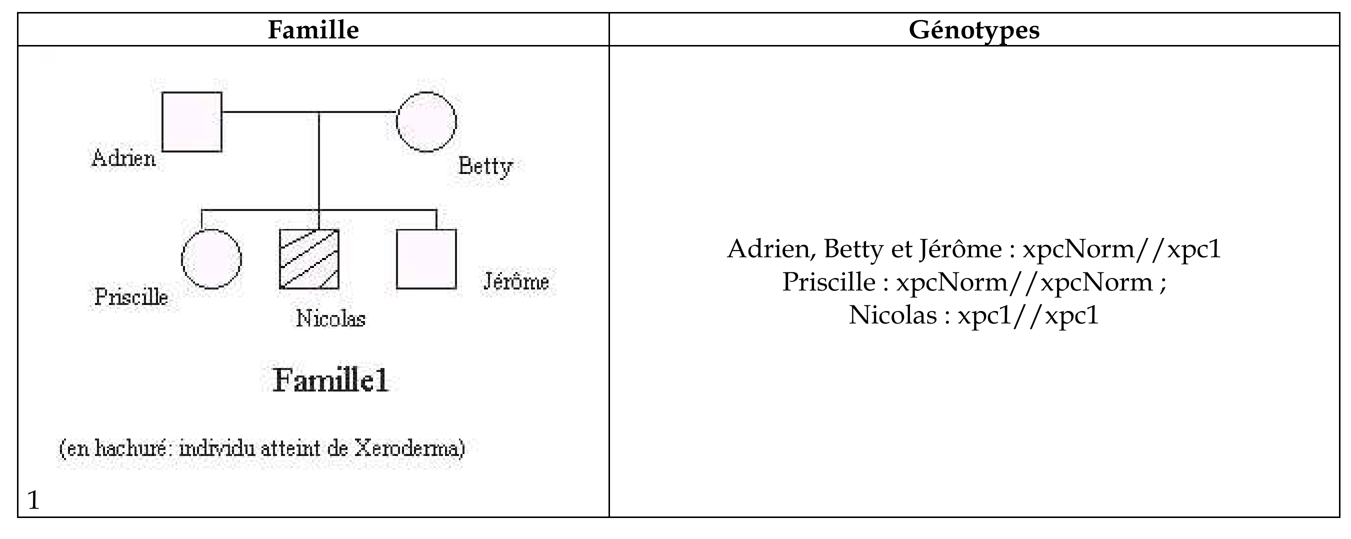
La commande **Xeroderma** de la banque de documents permet d'accéder aux fichiers :

- XeroarbreFamille1.bmp qui présente l'arbre généalogique de la famille 1 dont l'un des enfants est atteint de xeroderma ;
- XeroarbreFamille2.bmp qui présente l'arbre généalogique de la famille 2 dont l'un des enfants est atteint de xeroderma ;
- XeroarbreFamille3.bmp qui présente l'arbre généalogique de la famille 3 dont deux des enfants sont atteints de xeroderma.

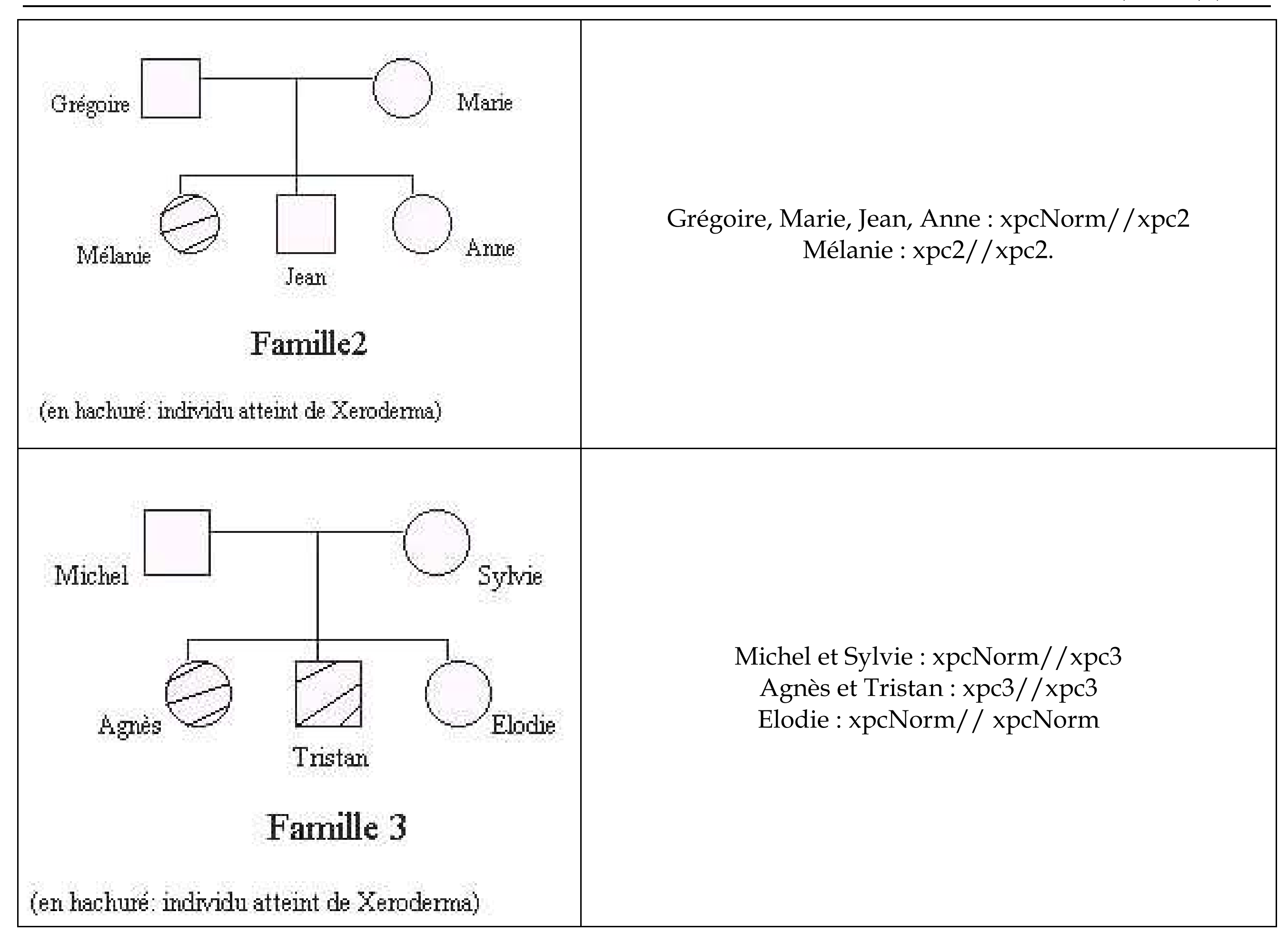
L'exploitation des données concernant les familles proposées permet de déterminer les relations de dominance/récessivité des allèles en cause, mais aussi de montrer qu'un phénotype xérodermique peut être dû à des mutations différentes.

Pour chaque famille, est fourni : l'arbre généalogique avec indication du phénotype clinique, les séquences des allèles de chaque individu pour le gène concerné, les allèles de référence du gène concerné.

La comparaison des allèles de chaque individu permet de déterminer son génotype. Les résultats figurent ci-dessous :



Famille	Génotypes
1	Adrien, Betty et Jérôme : xpcNorm//xpc1 Priscille : xpcNorm//xpcNorm ; Nicolas : xpc1//xpc1



La comparaison du génotype d'un individu hétérozygote avec son phénotype sain permet de dire que les allèles mutés sont récessifs par rapport à l'allèle normal. Ces relations de dominance/récessivité ne se conçoivent qu'au niveau du phénotype cellulaire et clinique, car tous les allèles s'expriment.

Des facteurs de l'environnement peuvent avoir une influence sur le phénotype

Documents fournis

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Xeroderma** permet d'accéder au fichier protect.bmp qui affiche un texte précisant les précautions que doivent prendre les individus xérodermiques pour éviter les UV.

L'influence de facteurs de l'environnement comme les UV peut être mise en évidence en comparant le phénotype de deux individus, ayant des génotypes identiques (par exemple xpc2//xpc2), mais qui vivent dans des conditions différentes. En effet, si l'un de ces individus est exposé aux rayons UV, il développera très rapidement des cancers cutanés ; par contre, l'individu qui prendra toutes les mesures de protection nécessaires vis-à-vis des UV ne développera pas de cancers cutanés.

Les possibilités d'édition de séquences offertes par le logiciel *Anagène* permettent à chaque enseignant de créer ses propres exemples.

Prédisposition génétique au cancer du sein

Informations scientifiques

(d'après un article de la Société française de génétique humaine, écrit par les docteurs D. Stoppa-Lyonnet, Institut Curie, C. Bonaïti-Pellie, INSERM, et le professeur Yves-Jean Bignon, unité d'oncogénétique – centre J.-Perrin).

Le cancer du sein touche environ, dans nos populations, une femme sur douze vivant jusqu'à l'âge de 80 ans. Avec environ 26 000 nouveaux cas et 11 000 décès chaque année, c'est la première cause de mortalité par cancer chez la femme.

Il existe plusieurs prédispositions au cancer du sein. Seule l'une d'entre elle sera étudiée ici.

Certaines familles sont très touchées par le cancer du sein ; dans ces familles, le risque d'être touché par ce cancer peut être multiplié par 8 à 10 par rapport à celui de la population générale.

Environ 10 à 15 % de ces cancers du sein sont liés à la transmission selon un mode autosomique dominant d'un allèle muté du gène BRAC1 ou du gène BRCA2 (la certitude qu'il s'agit d'une forme monogénique n'est obtenue que lorsqu'une mutation est identifiée).

Ce mode de transmission peut être évoqué lorsque l'on constate l'association des critères suivants :

- nombre de cas élevé dans une même famille ;
- précocité des atteintes cancéreuses (moins de 40 ans) ;
- atteinte mammaire bilatérale ;
- association fréquente avec un cancer de l'ovaire ;
- observation de cancer du sein chez l'Homme (rare).

Chez les femmes ayant hérité d'une mutation dominante de ce gène, le risque tumoral mammaire est élevé : environ 38 % à 50 ans ; la pénétrance est donc élevée, mais pas totale.

Les gènes BRCA1 et BRCA2 ont été identifiés en 1994 et 1995, et sont localisés respectivement sur les chromosomes 17 (17q21) et 13 (13q14). Ces gènes sont de grande taille : 5 592 nucléotides répartis en 22 exons pour la séquence codante du gène BRCA1 ; 10 254 nucléotides répartis en 26 exons pour le gène BRCA2 ; les exons 11 de chacun de ces gènes sont de très grande taille.

Les gènes BRCA1 et BRCA2 font partie des gènes suppresseurs de tumeur : les protéines codées par ces gènes sont localisées dans le noyau et sont impliquées dans la réparation de l'ADN, mais leur rôle précis n'est pas encore bien connu.

Les gènes BRCA1 et BRCA2 ne sont pas impliqués dans les formes de cancer non héréditaires.

On connaît de très nombreuses mutations de ces gènes BRCA1 et BRCA2 (plus de 900 mutations différentes connues au début de l'année 2000) ; la plupart de ces mutations sont de type « inactivateur », c'est-à-dire qu'elles entraînent la synthèse d'une protéine non fonctionnelle. On arrive à identifier la mutation en cause dans une famille (s'il y a implication du gène BRAC1 ou du gène BRCA2) dans environ 70 % des cas, en utilisant les techniques classiques (SSCP, DGGE, HA). D'autre part, on constate que la plupart des cellules tumorales ont perdu l'allèle normal.

On estime la fréquence des sujets porteurs de mutations BRCA1 entre 1/500 et 1/2000, et celle des sujets porteurs de mutations BRCA2 entre 1/2150 et 1/700.

Chez les sujets atteints, un test génétique peut être réalisé pour tenter d'identifier la mutation en cause ; cette recherche est très longue (plusieurs mois) car elle nécessite le criblage de l'ensemble des deux gènes. Si une mutation a été identifiée, un diagnostic présymptomatique peut être réalisé chez les personnes apparentées ; il s'agit d'un test simple, réalisable en quelques jours puisque la mutation a été identifiée. Si le résultat est négatif, la surveillance sera celle proposée habituellement à la population ; si le résultat est positif, une prise en charge adaptée est proposée avec une surveillance accrue (mammographie annuelle dès l'âge de 30 ans, échographie annuelle pour dépister un cancer de l'ovaire à partir de 35 ans…).

Des facteurs environnementaux peuvent favoriser la survenue des cancers du sein chez les individus porteurs d'un allèle muté : on a estimé que le risque tumoral mammaire chez une femme porteuse d'une mutation BRCA1 est près de trois fois plus important si elle est née après 1930 qu'avant. Il peut s'agir de facteurs alimentaires, qui influencent l'horloge hormonale.

Pistes d'exploitation pédagogiques des données fournies

Les séquences et documents fournis ici permettent d'aborder, en première S, la notion de prédisposition familiale et de gènes de susceptibilité dans le cas d'un cancer : le cancer du sein. Cette étude sera l'occasion de mettre l'accent sur l'intérêt du dépistage et de la surveillance accrue des personnes à risque.

Phénotype clinique et moléculaire en relation avec le génotype

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin Complexité **des relations génotype-phénotype-environnement/Cancer du sein** permet, par développement de l'arborescence, d'accéder à :

- **Allèles BRCA1** qui charge le fichier BCRA1.edi affichant les séquences strictement codantes de l'allèle de référence (allèle référence BRCA1) et de trois allèles mutés (allèle m1 BRCA1 – allèle m2 BRCA2 – allèle m3 BRCA3) du gène BRCA1, et des protéines correspondantes. Tous les allèles mutés fournis ici codent pour des protéines non fonctionnelles ;
- **Allèles BRCA2** qui charge le fichier BRCA2.edi affichant les séquences strictement codantes de l'allèle de référence (allèle référence BRCA1) et d'un allèle muté (allèle m1 BRCA2) du gène BRCA2, et les protéines correspondantes. L'allèle muté fourni ici code pour une protéine non fonctionnelle ;
- **Famille1 et allèles** qui charge le fichier BRCA1-famille1.edi affichant les allèles du gène BRCA1 des membres de la famille 1 dont l'arbre généalogique (fichier Famille1BRCA.bmp) est chargé simultanément. Les génotypes des membres de la famille sont les suivants :

(allèle ref. BRCA1//allèle ref. BRCA1) : I2, II1, II6 ; III1

(allele ref. BRCA1//allele mut2 BRCA1) : I1 ; II2 ; II3 ; II4 ; II5 ; III2 ; III3

Documents fournis

Dans la banque de thèmes d'étude, le chemin **Complexité des relations génotype-phénotype-environnement/Cancer du sein** permet d'atteindre **Gènes de susceptibilité** pour charger les fichiers :

- genes.bmp qui affiche un texte de présentation des gènes BRCA1 et BRCA2 ;
- courbe.bmp qui affiche la courbe indiquant la fréquence d'apparition d'un cancer du sein en fonction de l'âge chez des femmes porteuses d'un allèle muté du gène BRCA1 ou du gène BRCA2 (d'après un article de *Science*, vol. 302, 24 octobre 2003) ;
- arbreFamille1Cancer.bmp qui présente l'arbre généalogique de la famille 1. Dans cette famille, plusieurs cas de cancers du sein ont été observés, touchant plusieurs générations. Ils sont liés à la présence d'un allèle muté du gène BRCA1 (l'allèle mut2), qui provient à l'origine du père I1.

Arbre de la famille 1

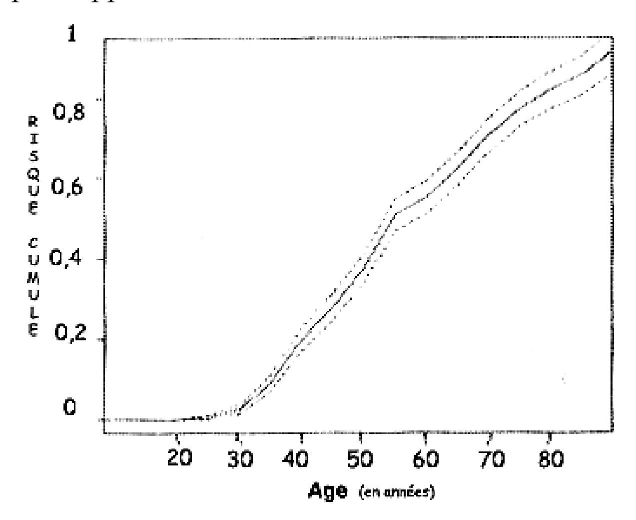
Homme non atteint de cancer

Femme non atteinte de cancer

Femme atteinte de cancer

II2 – atteinte d'un cancer du sein à 38 ans
II4 – atteinte d'un cancer du sein à 32 ans, puis d'un cancer des ovaires à 36 ans
II6 – atteinte d'un cancer du sein à 52 ans
III2 – atteinte d'un cancer du sein à 26 ans

Le constat fait sur l'arbre généalogique (nombre élevé de femmes atteintes d'un cancer du sein dans cette famille, âge précoce d'apparition du cancer) suggère l'implication d'un facteur héréditaire. Le texte de présentation des gènes BRCA et la courbe confirment que des mutations de certains gènes, comme BRCA1 et BRCA2 peuvent être mises en relation avec les phénotypes cancéreux étudiés, la présence d'un allèle muté augmentant beaucoup le risque d'apparition d'un cancer du sein.

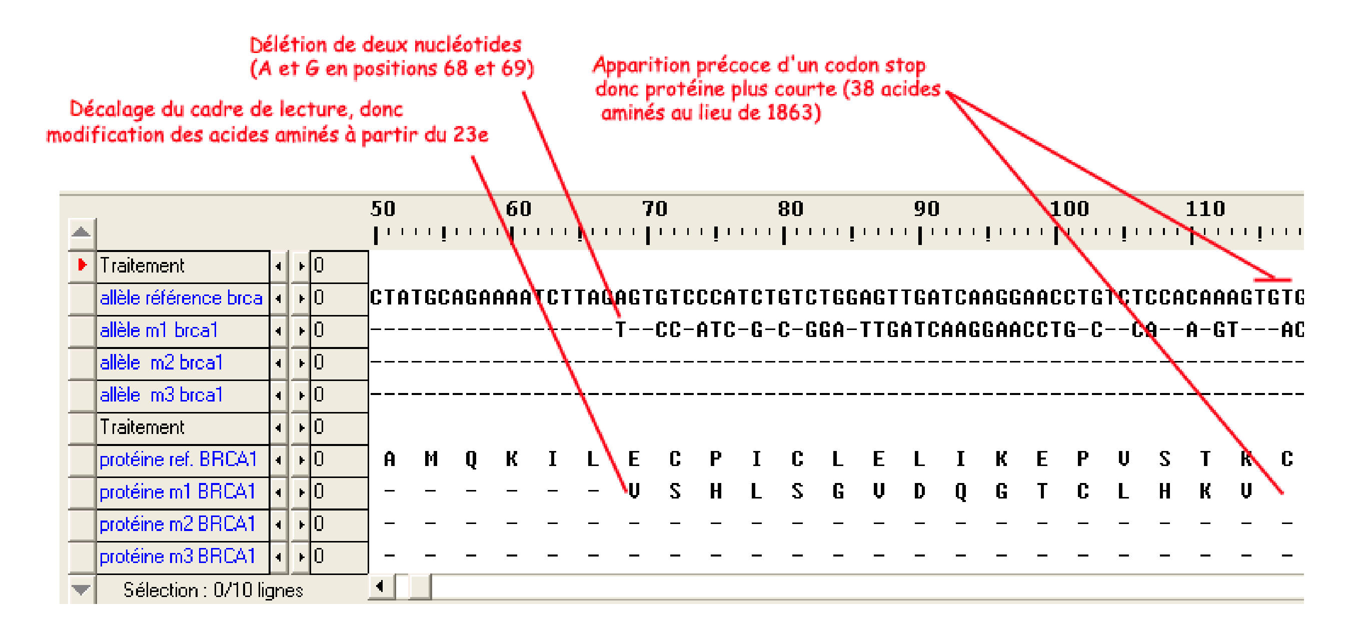


Risque d'apparition d'un cancer du sein chez les femmes porteuses d'un allèle muté du gène BRCA1 ou BRCA2 (d'après un article de Science*, vol. 302, 24 octobre 2003)*

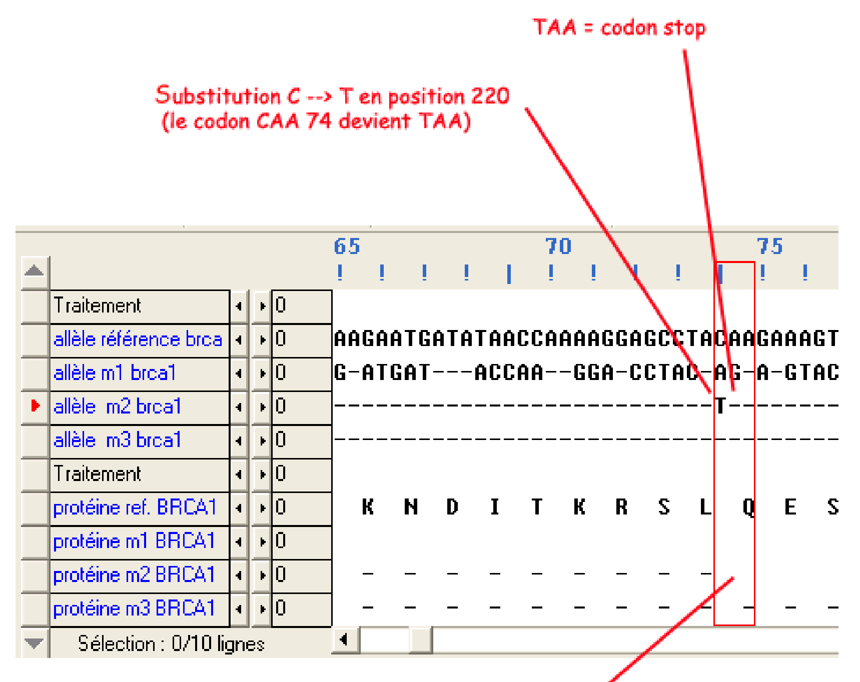
Le travail réalisé à partir des séquences fournies permet de relier génotype/phénotype moléculaire et phénotype clinique.

Comparaison simple des allèles du gène BRCA1 et mise en relation avec la comparaison simple des protéines BRCA1 correspondantes (les séquences nucléiques étant très longues, l'alignement avec discontinuité est déconseillé avec des machines peu performantes)

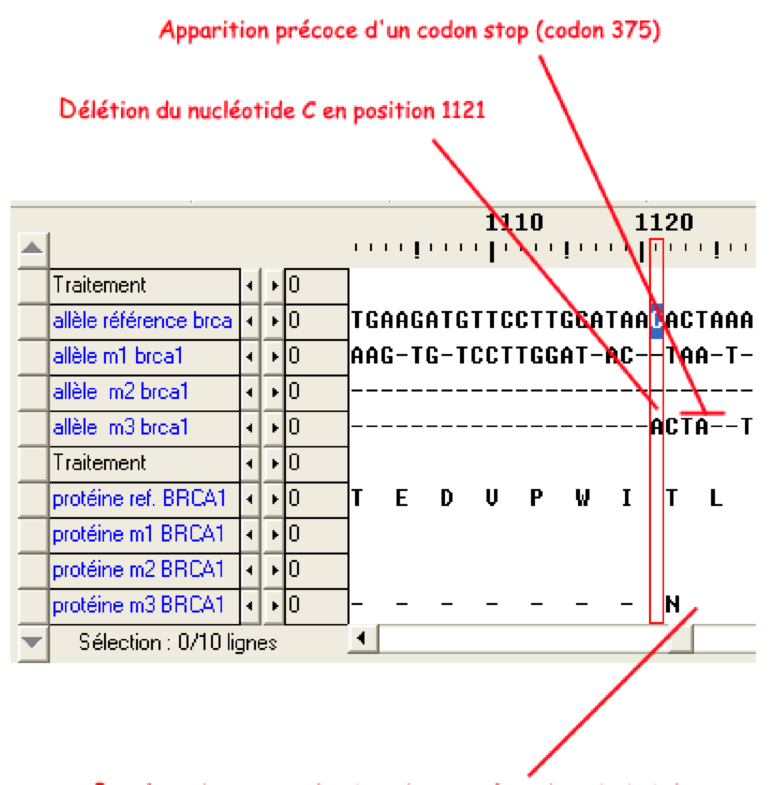
Comparaison Ref et m1



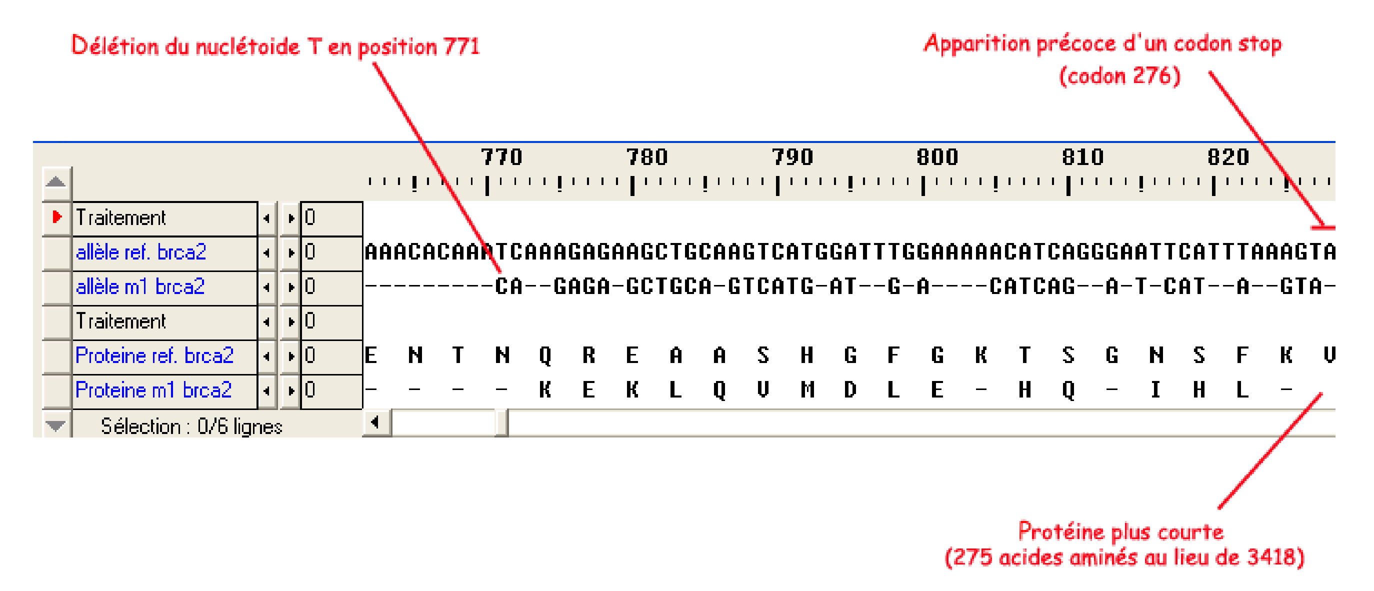
Comparaison Ref et m2



Comparaison Ref et m3



Comparaison simple des allèles du gène BRCA2 et mise en relation avec la comparaison simple des protéines BRCA2 correspondantes :

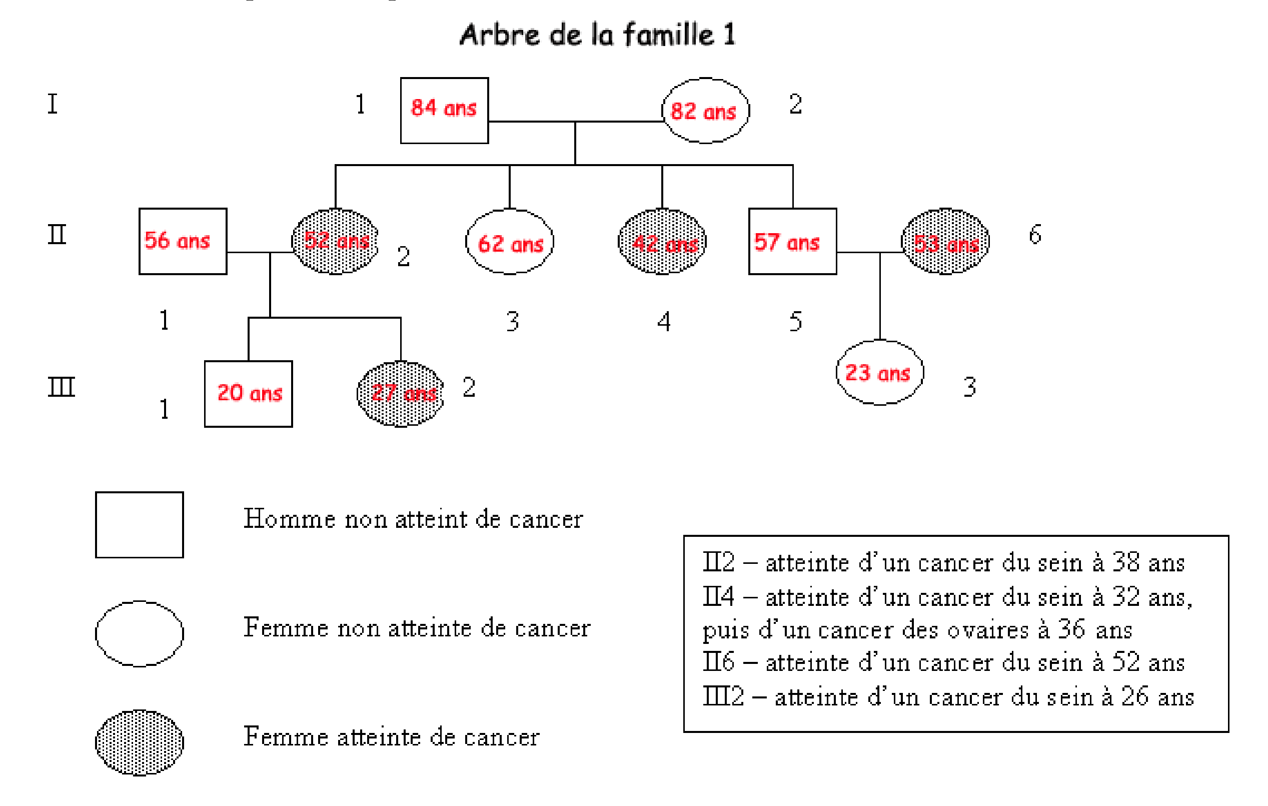


Conclusions

Les mutations présentes dans les allèles mutés fournis ici entraînent toutes la synthèse d'une protéine non fonctionnelle, le plus souvent tronquée à cause de l'apparition précoce d'un codon stop. Ces protéines ne peuvent donc remplir leur rôle réparateur de l'ADN dans la cellule, ce qui favorise l'apparition du cancer du sein.

Notion de prédisposition familiale

L'arbre généalogique suggère une prédisposition des femmes de cette famille au cancer du sein. Une des façons d'introduire le problème est de se demander si la femme III3 a un risque particulier d'avoir un cancer du sein, étant donné que beaucoup de femmes de sa famille en ont été atteintes.



La comparaison des allèles des membres de la famille avec les allèles du gène BRCA1 permet :

- d'identifier la mutation en cause dans cette famille : mutation m2 ;
- de montrer que l'allèle m2 a été apporté dans cette famille par le père I1 ;
- de déterminer le génotype de chacun des individus :
  - (allèle ref. BRCA1//allèle ref. BRCA1) : I2, II1, II6, III1 ;
  - (allèle ref. BRCA1//allele mut2 BRCA1) : I1, II2, II3, II4, II5, III2, III3 ;
- de mettre en évidence le fait que les femmes malades sont hétérozygotes, donc qu'un seul allèle suffit pour augmenter le risque de la maladie ;
- de constater que la femme II3, bien qu'hétérozygote, n'a pas développé de cancer du sein, ce qui permet de dire que le fait de posséder un allèle muté du gène BRCA1 n'implique pas forcément la maladie ;
- de dire que la femme III3 est hétérozygote et qu'elle a donc un risque supérieur à la normale de développer un cancer du sein.

On peut donc conclure que la possession d'un allèle muté du gène BRCA1 prédispose à l'apparition d'un cancer et que cette prévalence n'est pas totale.

D'autre part, il faut bien que l'élève comprenne l'intérêt d'un dépistage de la présence d'un allèle muté du gène BRCA1 (ou 2) : il s'agit de pouvoir dire à une femme si elle a un risque particulier et de lui proposer alors une surveillance accrue de façon à pouvoir intervenir précocement et à limiter les conséquences de la maladie.

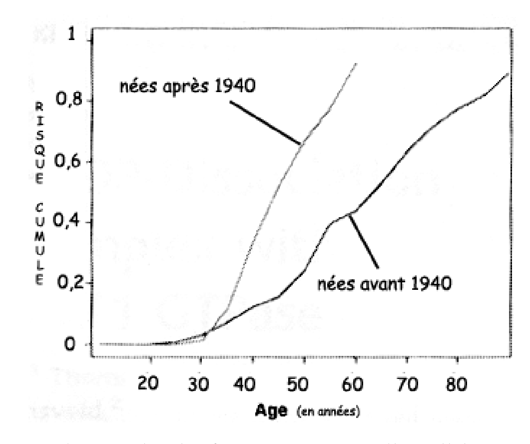
Influence d'autres facteurs

Documents

Dans la banque de documents, le chemin **Complexité des relations génotype-phénotype-environnement/Cancer du sein** permet d'atteindre **Influence de l'environnement** pour charger les fichiers :

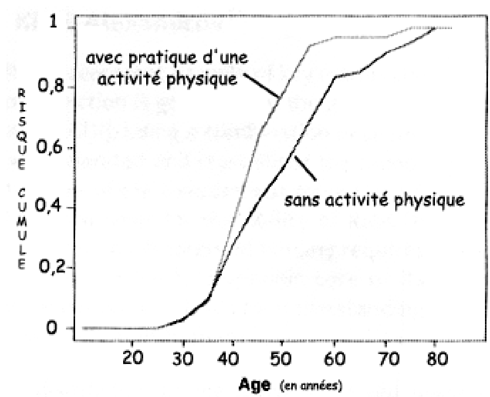
- preventionCancer.bmp qui affiche un texte présentant les mesures de prévention en faveur des personnes à risque particulier pour le cancer du sein ;
- age.bmp qui affiche une courbe indiquant la fréquence d'apparition d'un cancer du sein (en fonction de l'âge) selon l'année de naissance, chez des femmes porteuses d'un allèle muté du gène BRCA1 ou du gène BRCA2 (d'après un article de *Science*, vol. 302, 24 octobre 2003) ;
- activite.bmp qui affiche une courbe indiquant la fréquence d'apparition d'un cancer du sein (en fonction de l'âge) selon la pratique d'une activité physique chez des femmes porteuses d'un allèle muté du gène BRCA1 ou du gène BRCA2 (d'après un article de *Science*, vol. 302, 24 octobre 2003).

Ces documents permettent de mettre en évidence l'influence d'autres facteurs sur le phénotype « cancer du sein ».



Risque d'apparition d'un cancer du sein chez les femmes porteuses d'un allèle muté du gène BRCA1 ou BRCA2 (d'après un article de Science*, vol. 302, 24 octobre 2003)*

Ainsi, le fait que les femmes porteuses d'un allèle muté BRCA1 aient un risque différent de développer un cancer du sein selon qu'elles sont nées avant ou après 1940 suggère l'influence d'un facteur de l'environnement, probablement lié à l'alimentation.



Risque d'apparition d'un cancer du sein chez les femmes porteuses d'un allèle muté du gène BRCA1 ou BRCA2 (d'après un article de Science*, vol. 302, 24 octobre 2003)*

Enfin, le fait que les femmes porteuses d'un allèle muté BRCA1 aient un risque différent de développer un cancer du sein selon qu'elles sont actives physiquement ou non met aussi en évidence l'influence de l'environnement.

Cette pluralité des facteurs de prédisposition à l'apparition d'un cancer est importante à dégager.

MORPHOGENÈSE VÉGÉTALE ET ÉTABLISSEMENT DU PHÉNOTYPE

Polymorphisme des gibbérellines

Informations scientifiques

Le phénotype nain chez quelques végétaux

Chez l'Épinard de l'espèce *Spinacia oleracea*, la plante se présente en rosette en périodes de jours courts, mais lors du passage en jours longs, on observe une croissance importante de la tige. Cette croissance, due essentiellement à une élongation des entrenœuds, est sous l'influence de la photopériode et nécessite des hormones de croissance, les gibbérellines.

Chez le Pois de l'espèce *Pisum sativum*, on connaît des mutants qui présentent un phénotype en rosette, dû à une croissance quasi nulle des entrenœuds. Plusieurs gènes sont mis en cause, codant notamment pour des enzymes intervenant dans la chaîne de biosynthèse des gibbérellines.

Le phénotype « nain », qui se traduit par un port en rosette, semble donc dû à une croissance très limitée des entrenœuds. Des hormones comme les gibbérellines, mais aussi un facteur de l'environnement (la photopériode) semblent contrôler cette croissance.

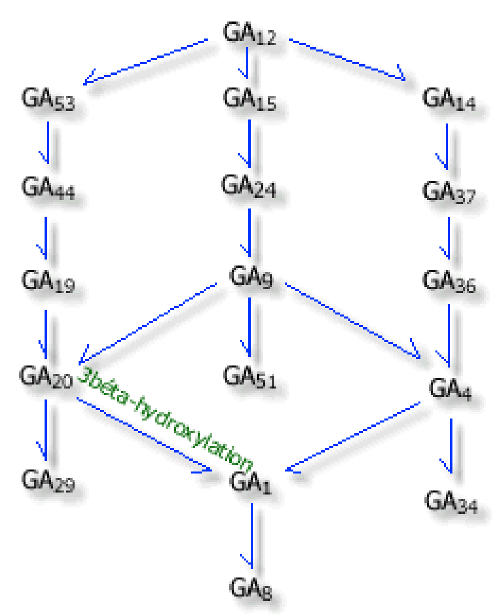
Les gibbérellines

Les gibbérellines sont produites dans les parties terminales des jeunes plantes (sauf dans les méristèmes), les pétioles, les jeunes feuilles.

Elles agissent notamment sur les cellules des entrenœuds en stimulant l'élongation cellulaire. Elles stimulent aussi la multiplication cellulaire au niveau des méristèmes intercalaires.

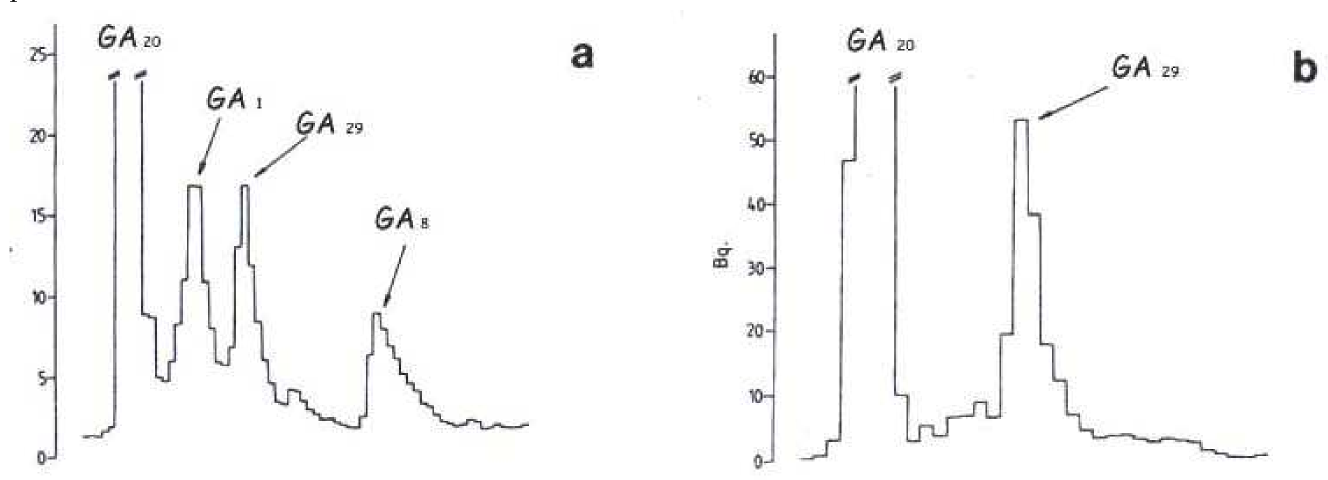
On connaît plus de cent gibbérellines différentes, qui sont toutes des diterpènes. Mais toutes n'ont pas la même activité et la même efficacité.

Le document ci-dessous présente les chaînes de biosynthèse des gibbérellines (chaque réaction est catalysée par une enzyme) :



Seule la gibbérelline GA1 est active. Les gibbérellines GA8, GA29 et GA34 résultent de la dégradation de leurs précurseurs.

On peut évaluer par une technique de chromatographie après marquage la quantité de quelques-unes des gibbérellines dans une plante. On peut ainsi comparer les gibbérellines présentes chez une plante de phénotype normal (a) et chez une plante de phénotype nain (b). Les résultats de ces chromatographies sont présentés ci-dessous :



Chez le Pois, le phénotype nain est donc dû à l'absence de transformation de la gibbérelline GA_{20} en gibbérelline GA_1.

Les mutations affectant la synthèse de gibbérellines

On ne s'intéressera ici qu'à la réaction de synthèse de la GA1 à partir de la GA20, catalysée par une enzyme « 3 béta-hydoxylase ». Cette enzyme est codée par un gène appelé « Le » chez le Pois. Ce gène s'exprime surtout dans la tige, la racine et les cotylédons du pois en croissance.

Chez le Pois, on connaît plusieurs allèles de ce gène, dont les deux allèles « Le » (codant pour une enzyme fonctionnelle) et « le » (codant pour une enzyme non fonctionnelle).

Les homozygotes le//le présentent un phénotype nain en rosette ; ce nanisme est plus ou moins accentué selon les allèles ; il est très important lorsque l'enzyme synthétisée est plus courte que la normale, la mutation ayant fait apparaître un codon stop anticipé.

On connaît une variété de Pois « ultra-nain » dont la taille ne dépasse pas 1 cm ; le gène muté dans ce cas n'est pas le gène « Le », mais le gène « Na », codant pour une enzyme intervenant plus en amont dans la chaîne de biosynthèse.

L'influence de l'environnement

Des mesures réalisées chez l'Épinard ont permis de montrer l'influence de la photopériode sur la croissance des entrenœuds.

Le tableau ci-dessous donne les concentrations de quelques gibbérellines lors du passage « jours courts/jours longs » chez l'Épinard (les valeurs sont données en ng/g de masse sèche) :

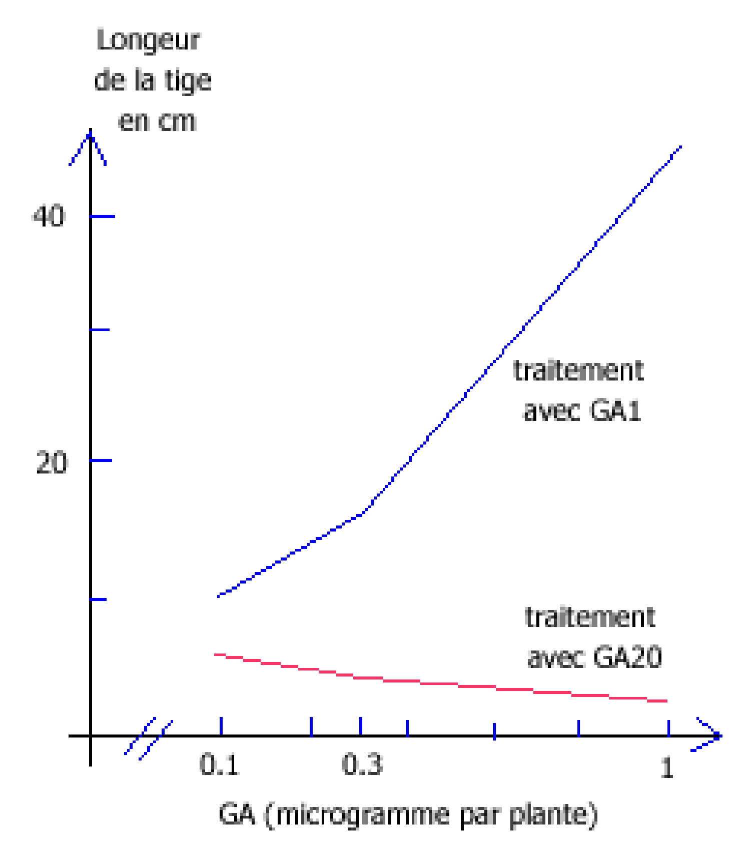
Gibbérellines	0 jour long	4 jours longs	8 jours longs	12 jours longs
G453	5,9	6,6	7,9	7,1
GA44	2,2	8,8	15,2	17,4
GA19	15	27,5	54,2	45,0
GA20	1,4	6,9	18,1	23,1
GA1	1,0	3,5	3,7	5,1

On constate donc que le passage en jours longs stimule toute la chaîne de biosynthèse de la gibbérelline GA8 à partir de la GA53.

On peut donc envisager l'influence d'un facteur de l'environnement sur le génome : il y aurait stimulation des gènes codant pour des enzymes impliquées dans cette chaîne de biosynthèse.

Des expériences réalisées chez l'Épinard ont permis de situer le niveau d'action de substances qui sont des inhibiteurs de croissance, comme le BX-112 par exemple. L'une de ces expériences consiste à appliquer cet inhibiteur de croissance à un plant d'Épinard de jours longs, puis à appliquer des doses croissantes de GA20 et de GA1 ; les mesures sont effectuées après 22 « jours longs ».

Le graphique ci-dessous présente les résultats obtenus :



L'inhibiteur de croissance BX-112 empêche donc la réaction de transformation de la GA20 en GA1. Il semble agir au niveau du génome, en limitant l'expression du gène codant pour la 3 béta-hydroxylase.

Pistes d'exploitation pédagogique des données fournies

Les données proposées permettent d'aborder les notions suivantes :

- le phénotype peut se définir à différentes échelles ;
- le génotype détermine le phénotype moléculaire, lui-même responsable du phénotype cellulaire et macroscopique ;
- le phénotype dépend du génotype, mais aussi de facteurs de l'environnement.

Les résultats et conclusions que l'on peut obtenir à l'aide de ces données sont présentés ci-après.

Le phénotype nain chez les végétaux peut être défini à plusieurs niveaux

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Morphogenèse végétale et établissement du phénotype/Phénotype végétal, génotype et environnement/Polymorphisme des gibbérellines : Polymorphisme gène LE** permet d'atteindre la commande **Allèles gène LE** qui charge le fichier alleles-LE.edi affichant les séquences strictement codantes de cinq allèles du gène de la 3 beta-hydroxylase ; l'allèle « Le » est l'allèle normal ; les allèles « le1 », « le2 », « le3 », « le4 » sont des allèles mutés codant pour des enzymes non fonctionnelles.

Documents fournis

Dans la banque de documents, le chemin **Morphogenèse végétale et établissement du phénotype/Phénotype végétal, génotype et environnement** permet d'atteindre **Phénotype nain et synthèse des gibbérellines** pour charger les fichiers :

- phenotypenain.bmp qui affiche un texte présentant les caractéristiques macroscopiques et cellulaires du phénotype nain chez les végétaux ; ce texte signale aussi l'importance des gibérellines pour une croissance normale de la plante ;

- biosynth-gibb.bmp par la commande **biosynthèse gibbérellines** affichant un schéma présentant différentes voies de la biosynthèse des gibbérellines ;

- chromato.bmp présentant les résultats d'une chromatographie comparative entre un Pois normal et un Pois nain. Cette chromatographie révèle les quantités relatives de quelques gibbérellines et permet de relier ainsi le phénotype macroscopique au phénotype biochimique.

	Phénotype macroscopique	**Phénotype cellulaire**	**Phénotype biochimique**	**Phénotype moléculaire**
Plante normale	Développement normal – entrenœuds développés	Élongation cellulaire normale – Multiplication cellulaire normale	Présence de GA1 en grande quantité	Enzyme 3 bêta-hydroxylase fonctionnelle
Plante naine	Plante en rosette – entrenœuds très courts	Élongation cellulaire très réduite et multiplication cellulaire limitée	Quantités de GA1 très réduites	Enzyme 3 bêta-hydroxylase inactive

La mise en relation de la comparaison des allèles du gène pour la 3 bêta-hydroxylase avec la comparaison des protéines correspondantes permet d'aboutir aux résultats et conclusions suivants :

<table>
<tr><th>Allèle</th><th>Changements dans la séquence nucléique</th><th>Changements dans la séquence protéique</th><th>Type de mutation</th><th>Conséquences</th></tr>
<tr><td>Allèle Le</td><td colspan="4">référence</td></tr>
<tr><td>Allèle le1</td><td>Del du G en 376</td><td>Modification à partir de l'acide aminé 126, et polypeptide plus court (172 acides aminés au lieu de 374)</td><td>Décalante</td><td rowspan="5">Enzyme non fonctionnelle, donc pas synthèse de GA1, donc entrenœuds plus courts et port en rosette</td></tr>
<tr><td>Allèle le2</td><td>G → A en 685 (GCC → ACC en position 229)</td><td>A → T en 229</td><td>Faux-sens</td></tr>
<tr><td>Allèle le3</td><td>C → T en 826 (CAT → TAT en 276)</td><td>H → Y en 276</td><td>Faux-sens</td></tr>
<tr><td rowspan="2">Allèle le4</td><td>A → T en 267 (TTA → TTT en 89)</td><td>L → F en 89</td><td>Faux-sens</td></tr>
<tr><td>C → T en 826 (CAT → TAT en 276)</td><td>H → Y en 276</td><td>Faux-sens</td></tr>
</table>

Le phénotype dépend aussi de facteurs de l'environnement

Documents

Dans la banque de documents, le chemin **Morphogenèse végétale et établissement du phénotype/Phénotype végétal, génotype et environnement** permet d'atteindre **Influence de l'environnement** pour charger les fichiers :

- expenvironnement.bmp qui affiche les résultats de mesures mettant en évidence l'influence de la photopériode sur la chaîne de biosynthèse des gibbérellines ;
- inhibcroissance.bmp qui présente les résultats d'une expérience menée chez l'Épinard et mettant en évidence l'influence d'une molécule chimique (un inhibiteur de croissance : le BX 112) sur la chaîne de biosynthèse des gibbérellines.

L'analyse des documents fournis met en évidence l'influence de facteur de l'environnement sur le génotype :

- l'allongement de la photopériode stimule l'expression de certains gènes, favorisant ainsi la chaîne de biosynthèse de la GA_1 ;
- des substances chimiques comme l'inhibiteur de croissance BX112 inhibe l'expression de certains gènes, ce qui réduit la production de GA_1 ; la plante possède alors un phénotype nain, alors qu'aucune mutation particulière n'affecte les gènes en cause.

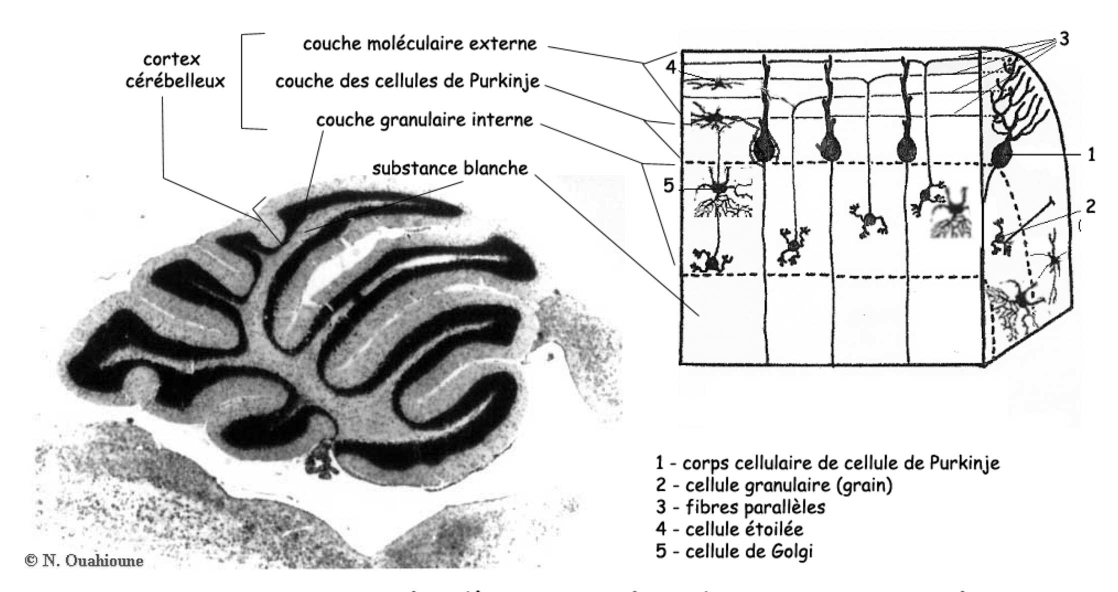
GÉNOTYPE, ENVIRONNEMENT ET FONCTIONNEMENT DU SYSTÈME NERVEUX

Les phénotypes « mutants cérébelleux »

Informations scientifiques

L'organisation du cortex cérébelleux

Le cortex cérébelleux est une petite structure relativement simple qui ne contient que cinq types cellulaires organisés en trois couches :



Coupe parasaggitale d'un cervelet de souris normale

- **la couche externe ou moléculaire**, qui contient les arborisations dendritiques des cellules de Purkinje, les axones des cellules olivaires ou fibres grimpantes, les axones des cellules granulaires ou fibres parallèles, ainsi que des interneurones (cellules étoilées et cellules en corbeille) ;
- **la couche moyenne ou couche des cellules de Purkinje** qui contient le corps des cellules de Purkinje répartis en une seule assise ;
- **la couche interne ou granulaire** dont les éléments dominants sont des corps cellulaires très nombreux et de petite taille, les grains ou cellules granulaires. Les axones de ces cellules remontent dans la couche moléculaire et se divisent en « T » en formant de longues branches, les fibres parallèles. Cette couche interne contient aussi des cellules de Golgi, localisées au voisinage du soma des cellules de Purkinje.

Le cortex cérébelleux est une structure hautement répétitive, constituée au niveau microscopique par la répétition un très grand nombre de fois du même circuit élémentaire. Ce circuit est organisé autour de la cellule de Purkinje.

Il existe essentiellement deux systèmes excitateurs : les fibres grimpantes et les fibres moussues ; il existe également des interneurones inhibiteurs : les cellules de Golgi, les cellules en panier et les cellules étoilées.

La cellule de Purkinje est l'élément central du réseau synaptique cortical cérébelleux : les interneurones de la couche moléculaire, cellules étoilées et en corbeille, ainsi que les cellules des grains de la couche granulaire interne se projettent majoritairement sur les cellules de Purkinje. Cette cellule de Purkinje a une forme spécifique, avec notamment un réseau dendritique très développé. L'axone, qui émerge de la région basale du corps cellulaire, traverse la couche granulaire interne, rejoint la substance blanche et se termine dans les noyaux profonds ou même en dehors du cervelet, au niveau des noyaux vestibulaires.

Des cellules gliales existent aussi dans le cervelet : les cellules macrogliales (astrocytes et oligodendrocytes), qui contribuent à la croissance et à la survie des neurones, et les cellules microgliales, qui constituent la principale population de macrophages cérébraux.

Les fonctions du cervelet

Le cervelet reçoit des afférences du cortex moteur, des récepteurs cutanés proprioceptifs, visuels, tactiles et auditifs, et des récepteurs viscéraux.

L'archéocervelet (lobe floculo-nodulaire) contrôle les mécanismes de la station érigée ; il intervient sur les mécanismes réflexes du redressement et de l'équilibration.

Le paléocervelet (vermis) contrôle la motilité axiale et la motilité d'attitude.

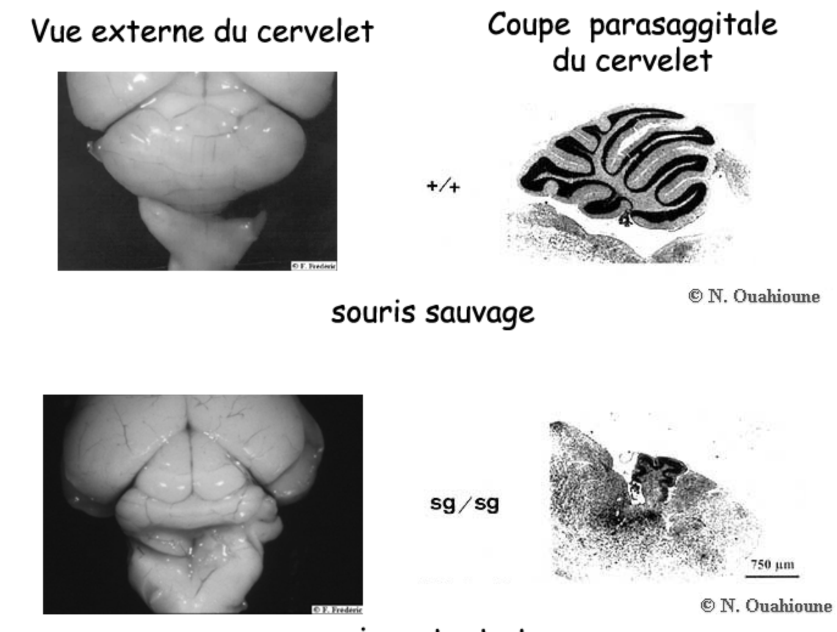
Le néocervelet (lobes latéraux) intervient dans la régulation du geste.

Les phénotypes mutants cérébelleux chez la Souris

On connaît plusieurs phénotypes « mutants cérébelleux » chez la Souris.

Les mutants « staggerer » (sg)

Ces mutants présentent un syndrome ataxique grave et survivent difficilement après le sevrage. Dès le 8e jour, le souriceau staggerer est très hésitant dans sa démarche, avec une tendance à tituber et à traîner les pattes. Il présente un cervelet atrophié, avec des cellules de Purkinje en nombre très réduit (60 à 90 % des cellules manquent à un mois), et avec un arbre dendritique réduit.



On observe chez ces mutants des anomalies biochimiques au cours du développement du cervelet, liées à un déficit de l'activité d'une neuramidase dans le cervelet, ainsi qu'une réduction de la concentration en thymidine kinase. Ces mutants présentent aussi une susceptibilité exagérée à l'athérosclérose, probablement liée à une baisse du HDL.

Cette mutation autosomale récessive d'un gène situé sur le chromosome 9 s'exprime directement dans les cellules de Purkinje. Le gène a été récemment cloné, et il code pour le facteur de transcription ROR-alpha, qui est un récepteur nucléaire appartenant à la même famille que les récepteurs à l'acide 9-cis-rétinoïque (récepteurs des hormones thyroïdiennes et stéroïdiennes…). La protéine ROR-alpha est fortement exprimée dans les cellules de Purkinje durant le développement postnatal du cervelet.

La mutation staggerer correspond à une délétion se produisant dans la partie du gène codant pour le site de liaison sur l'ADN. L'absence de fonctionnalité de la protéine ROR-alpha pourrait modifier les interactions avec d'autres protéines, notamment les récepteurs nucléaires des hormones thyroïdiennes, qui jouent un rôle aussi bien dans le développement du système nerveux que dans la différenciation des cellules de Purkinje. Compte tenu du rôle central de cette cellule dans le réseau synaptique cortical du cortex cérébelleux, on comprend que son absence ou l'atrophie de son arbre dendritique puisse être à l'origine d'un développement anormal du cervelet, d'où le comportement ataxique observé.

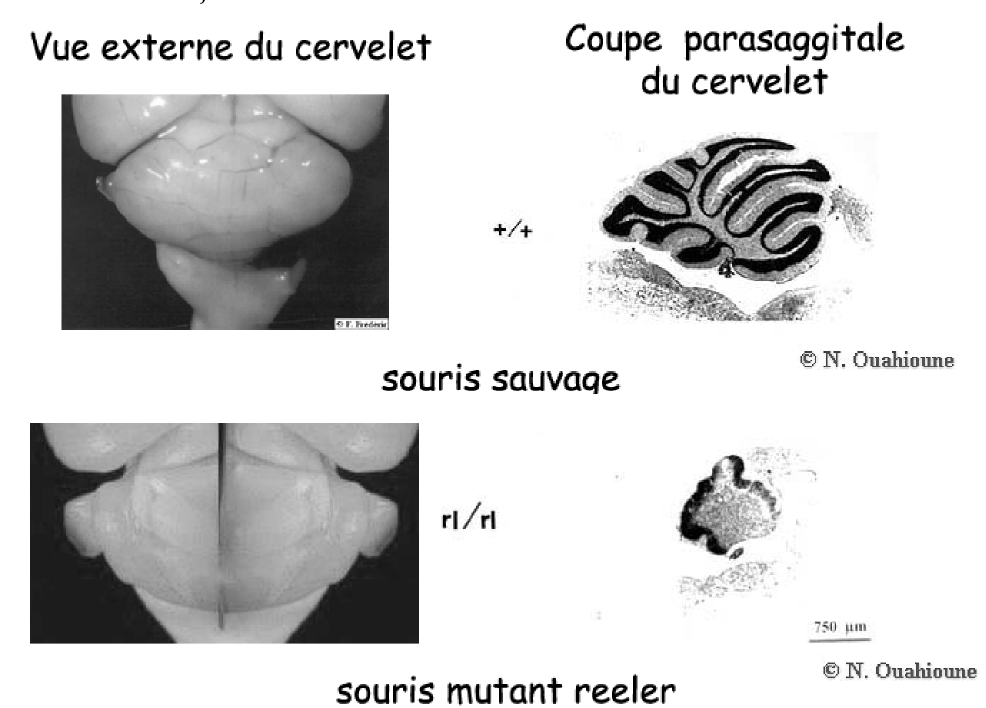
Chez l'hétérozygote, le développement du cervelet est quasi normal, mais on observe tout au long de la vie une perte progressive et partielle des cellules de Purkinje et une réduction de leur arbre dendritique.

Les mutants « reeler » (rl)

Ces mutants présentent un syndrome ataxique grave : ils titubent, chancellent. Les mouvements n'atteignent pas leur but, bien que leur force soit normale. L'animal est tremblant et a un comportement d'évitement pour réaliser certaines tâches.

Ce phénotype est lié à une mutation d'un gène situé sur le chromosome 5 ; cette mutation, autosomale récessive, affecte le développement embryonnaire du cerveau, et atteint surtout les structures laminées comme le cervelet, le cortex cérébral et l'hippocampe.

Chez l'embryon de Souris mutante, la migration des neurones se déroule normalement jusqu'au moment où ces neurones arrivent à destination. L'organisation normale en couches ne se fait alors pas correctement ; la disposition des cellules de Purkinje et l'orientation des neurones pyramidaux deviennent aléatoires, et le cervelet apparaît désorganisé et constitué de l'emboîtement de deux couches (un cortex cérébelleux d'architecture normale, mais atrophié, et à l'intérieur une masse cellulaire comportant la plupart des cellules de Purkinje mélangées aux cellules des noyaux profonds). Dès les premières semaines, on observe une perte de 50 % des cellules de Purkinje.



Le gène en cause a récemment été cloné ; il s'exprime très tôt au cours du développement et code pour une protéine « reelin » qui est une grosse protéine (400 kDa) de la matrice extacellulaire. Une des mutations consiste en une délétion d'un exon, ce qui aboutit probablement à une protéine tronquée dans sa partie C-terminale ; la protéine est alors retenue dans le RER où elle s'accumule.

Le développement du cervelet est normal chez l'hétérozygote.

Ces mutants constituent de bons modèles pour étudier l'implication du génotype dans la mise en place des réseaux neuroniques et donc du fonctionnement du système nerveux.

Pistes d'exploitation pédagogiques des données fournies

Les données proposées sur les phénotypes mutants cérébelleux chez la Souris permettent d'aborder les notions suivantes :

- le phénotype peut se définir à différentes échelles ;
- le génotype détermine le phénotype moléculaire, lui-même responsable du phénotype clinique ;
- un même phénotype clinique peut correspondre à plusieurs génotypes.

Cet exemple peut être utilisé pour bâtir un exercice permettant aux élèves de mobiliser les connaissances et méthodes de travail acquises à propos de l'étude d'autres exemples de phénotypes alternatifs.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Génotype, environnement et fonctionnement du système nerveux/Les phénotypes mutants cérebelleux staggerer et reeler** permet d'atteindre :

- **Polymorphisme gène Rora/mutation staggerer** qui charge le fichier mutationStaggerer.edi affichant les séquences nucléiques strictement codantes de l'allèle normal (allèleRORa Norm) et de l'allèle muté staggerer (allèleRORa staggerer) du gène codant pour la protéine ROR alpha ; séquences protéiques correspondantes (ProtRORa Norm et ProtRORa staggerer) ;

- **Polymorphisme gène Reelin/mutation reeler** qui charge le fichier mutationReeler.edi affichant les séquences nucléiques strictement codantes de l'allèle normale (allèle reelin normal) et de l'allèle muté reeler (allèle reeler) ; séquences protéiques correspondantes.

Documents fournis

Le développement de l'arborescence de **Génotype, environnement et fonctionnement du système nerveux** de la banque de documents permet d'atteindre **Les phénotypes mutants cérebelleux staggerer et reeler** pour charger les fichiers :

- cervelet.jpg qui présente la dissection d'un Souris normale montrant les centres nerveux et permettant de localiser le cervelet ; la photo légendée est accompagnée d'un petit texte précisant le rôle du cervelet ;
- coupecervelet.jpg qui présente la coupe de cervelet (coloration à la thionine phéniquée) permettant de voir son organisation et schéma de l'organisation tissulaire du cortex cérébelleux ; la photo et le schéma sont accompagnés d'un petit texte insistant sur les points importants de l'organisation du cortex cérébelleux ;
- staggerercervelet.bmp qui montre la comparaison des cervelets de Souris normale et de Souris mutante staggerer en vue externe et en coupe parasaggitale (coloration à la thionine phéniquée) ;
- reelercervelet.bmp qui montre la comparaison des cervelets de Souris normale et de Souris mutante reeler en vue externe et en coupe parasaggitale (coloration à la thionine phéniquée) ;
- phenotypestaggerer.bmp qui affiche un texte de description du phénotype des mutants staggerer ;
- phenotypereeler.bmp qui affiche un texte de description du phénotype des mutants reeler ;
- geneStaggerer.bmp qui affiche un texte présentant le gène mis en cause chez les mutants staggerer ;
- geneReeler.bmp qui affiche un texte présentant le gène mis en cause chez les mutants reeler.

NB : Pour comparer les allèles dont la différence de taille est importante, il est préférable d'utiliser le mode de comparaison « alignement avec discontinuités ».

Les phénotypes peuvent être définis à plusieurs niveaux

	Phénotype mutant staggerer	**Phénotype mutant reeler**
Phénotype clinique	Ataxie grave – cervelet de taille nettement inférieure à la normale	Ataxie grave – cervelet de taille nettement inférieure à la moyenne
Phénotype cellulaire	Nombre très réduit de cellules de Purkinje, celles qui sont présentes ayant un arbre dendritique très réduit. Désorganisation du cervelet.	Perte de 50 % des cellules de Purkinje dès le premier mois. Désorganisation du cervelet.
Phénotype moléculaire	Protéine ROR alpha (facteur de transcription) anormale	Protéine Reelin anormale

Déterminisme génétique chez le mutant staggerer

L'allèle RORa staggerer diffère beaucoup de l'allèle RORa Norm : il ne possède que 1 452 nucléotides au lieu de 1 572 chez l'allèle RORa Norm. La mutation consiste donc en une grosse délétion de toute une séquence de 120 nucléotides. Cet allèle muté code donc pour une protéine RORa tronquée donc non fonctionnelle. Cette protéine étant indispensable à la mise en place et à l'organisation du cortex cérebelleux, et surtout au bon développement des cellules de Purkinje (cellules jouant un rôle fondamental dans le cervelet), le cervelet de ces mutants sera désorganisé. Cet organe ayant un rôle important dans le maintien de la posture et le contrôle des mouvements, les Souris mutantes staggerer auront donc des mouvements désordonnés, et présenteront une ataxie importante.

Déterminisme génétique chez le mutant reeler

L'allèle reeler diffère beaucoup de l'allèle reelin normal : il est beaucoup plus court. Cet allèle code donc pour une protéine reelin tronquée donc non fonctionnelle. Cette protéine étant indispensable à la bonne disposition des cellules lors du développement du cervelet, le cervelet de ces mutants sera désorganisé. Cet organe ayant un rôle important dans le maintien de la posture et le contrôle des mouvements, les Souris mutantes staggerer auront donc des mouvements désordonnés, et présenteront une ataxie importante.

Un même phénotype clinique peut correspondre à plusieurs génotypes

Les Souris mutantes staggerer et reeler présentent pratiquement les mêmes symptômes (ataxie sévère) et ont toutes deux un cervelet de taille réduite et mal organisé. Pourtant, leur génotype est différent, puisque ce ne sont pas les mêmes gènes qui sont en cause dans les deux cas (gène RORa pour les mutants staggerer et gène Reelin pour les mutants reeler).

Ces données apportent aussi des arguments en faveur d'une influence du génotype sur la mise en place des réseaux de neurones et le fonctionnement du système nerveux, et peuvent donc être aussi utilisées dans la partie du programme intitulée « La part du génotype et la part de l'expérience individuelle dans le fonctionnement du système nerveux ».

SUGGESTIONS PÉDAGOGIQUES : CLASSE TERMINALE SÉRIE S

INTRODUCTION : LA BANQUE DE DONNÉES ET LES NOTIONS DU PROGRAMME

Les séquences et documents réunis peuvent être utilisés pour construire les notions des diverses parties du programme de biologie de terminale S. Certaines de ces séquences sont également fournies dans le logiciel *Phylogène* et sont susceptibles d'être traitées à l'aide de fonctions propres à ce logiciel, comme celle de construction d'arbre par exemple. Les deux logiciels sont donc complémentaires.

1 - Pour la partie « **Parenté entre êtres vivants actuels et fossiles** », le choix des molécules permet de rechercher les relations de parenté au sein d'ensembles plus ou moins étendus. Ainsi, le gène **CDC2** qui code pour un polypeptide jouant un rôle majeur dans la réalisation du cycle cellulaire peut être utilisé pour les **relations de parenté au sein des eucaryotes**. Les **gènes homéotiques** permettent de rechercher les parentés chez les **métazoaires bilatéraliens**. Les gènes codant pour les **globines alpha et bêta** se prêtent à l'établissement de **phylogénies chez les Vertébrés** et le gène codant pour le **pigment visuel** de courte longueur d'onde (opsine S ou « bleue ») peut être utilisé pour préciser les relations de parenté chez les **Primates**. Enfin, la banque contient des séquences de **l'ADN mitochondrial** de Primates actuels et de l'Homme de Neandertal. Ces séquences peuvent servir pour préciser les relations de parenté chez les **Primates** mais surtout avec l'exploitation de l'ADN fossile des neandertaliens permettent de discuter des relations entre **l'Homme de Neandertal et Homo sapiens**.

2 - En ce qui concerne la partie « **Stabilité et variabilité des génomes et évolution** », les séquences et documents fournis permettent d'envisager les différents types d'innovations génétiques et les mécanismes susceptibles d'assurer leur maintien dans les populations.

- L'exemple des **globines**, classique mais très documenté, est le plus riche et peut être utilisé pour établir les principales notions. En effet, l'analyse des divers allèles du gène bêta conduit à faire le point sur les différents types de **mutations ponctuelles** et celle des gènes codant pour les différentes chaînes de globine permet d'établir la notion de **famille multigénique** et donc celle de production de nouveaux gènes par duplication. La comparaison des séquences du gène alpha ou bêta chez différentes espèces ou celle des différents gènes de la famille multigénique dans l'espèce humaine mettent en évidence des sites conservés et des sites à forte variabilité. L'interprétation de ces sites à l'aide des documents fournis conduit aux idées de **conservation de mutations neutres** (par dérive génique) et **d'élimination des mutations affectant la fonction des globines**. Enfin, l'exemple classique de l'avantage procuré par la possession de l'allèle HbS chez les hétérozygotes en région paludéenne, permet de comprendre la **permanence et l'extension d'un allèle par sélection naturelle** dans certaines conditions d'environnement.
- Les exemples des gènes codant pour **l'alpha-antitrypsine** et la **G6PD** permettent aussi d'établir les différents types de **mutations ponctuelles** et surtout d'établir des **filiations entre allèles** et donc de reconstruire au moins en partie l'histoire d'un gène au sein des populations humaines. En outre, grâce à la riche documentation associée, l'exemple de la G6PD renforce celui de l'hémoglobine S pour faire saisir comment dans certaines conditions d'environnement, certains allèles se répandent par **sélection positive** dans certaines populations.
- L'importance du processus de **duplication suivi de mutations ponctuelles** des gènes dupliqués peut être appréhendée à partir de divers exemples de familles multigéniques : celle des **opsines** chez les Primates (à l'origine de la possibilité de vision des couleurs), celle des hormones hypophysaires LH, FSH, TSH et placentaire HCG, celle des hormones hypophysaires Gh (hormone de croissance), HPRL (prolactine) et HLP (hormone lactogène placentaire).
- Enfin, un des exemples les plus remarquables associant innovations génétiques et sélection naturelle est celui de **l'acquisition de la résistance aux insecticides par les Moustiques**. Sans traiter complètement le problème, on y voit comment des duplications de gènes codant pour les estérases, enzymes capables de dégrader les insecticides organophosphorés, arrivent à se répandre dans les régions où l'épandage de ces insecticides a été important de sorte qu'une grande partie de la population de Moustiques est devenue résistante.

3 - Pour la partie « **Procréation** », les documents fournis sont relatifs à **des cas cliniques d'infertilité** due à des allèles mutés de gènes intervenant soit dans la différenciation sexuelle au cours de la vie embryonnaire ou fœtale, soit dans la commande hormonale de l'appareil génital à la puberté. Les données morphologiques, anatomiques et les tests biologiques doivent permettre de faire des hypothèses sur le gène en cause dans les différents cas cliniques. Les séquences des allèles des différents gènes présents chez l'individu comparées à celles d'allèles fonctionnels de référence permettent de tester les hypothèses émises.

4 - Pour la partie « **Immunologie** », les séquences fournies sont celles des chaînes lourdes et légères **d'immunoglobulines anti VIH**. La comparaison des séquences de ces chaînes permet de dégager la notion de partie constante et de partie variable de chaque type de chaîne et donc de faire l'hypothèse que la spécificité des anticorps est liée aux régions variables (et plus précisément aux zones hypervariables). Cela peut être testé à l'aide du logiciel *Rasmol* ou *RasTop* qui permet de localiser sur la structure spatiale d'une immunoglobuline les sites de fixation antigénique et de constater qu'ils correspondent aux régions hypervariables des chaînes d'immunoglobulines. Un travail du même type peut être conduit à partir des séquences des récepteurs des lymphocytes T.

5 - En ce qui concerne **l'Enseignement de spécialité**, la banque de données fournit des séquences relatives à une famille où des individus sont **albinos**. L'exemple est relativement complexe dans la mesure où l'albinisme dans l'arbre généalogique fourni a pour origine deux gènes différents, mais il est possible d'avoir une démarche très progressive. Surtout, les données fournies permettent de saisir comment agissent les enzymes de restriction et comment les fragments d'ADN obtenus par l'action de ces enzymes peuvent être séparés par électrophorèse.

PARENTÉ ENTRE ÊTRES VIVANTS ACTUELS ET FOSSILES – PHYLOGENÈSE, ÉVOLUTION

Relations de parenté au sein du vivant – Gène CDC2

Informations scientifiques

Des informations complémentaires sont disponibles sur le site bmedia :

http://www.snv.jussieu.fr/bmedia/cyclecellBM/index.htm

Le cycle cellulaire

Un cycle cellulaire constitué d'une interphase et d'une mitose est observable chez tous les eucaryotes. L'interphase comprend elle-même trois phases : la phase G1 (ou Gap 1), la phase S (réplication de l'ADN) et la phase G2 (ou Gap 2). La durée totale d'un cycle cellulaire varie beaucoup d'une cellule à l'autre. La durée de chacune des phases du cycle cellulaire peut être calculée si l'on connaît le pourcentage de cellules en phases S (incorporation de BrdU), G1 et G2 (estimation de la quantité moyenne d'ADN par noyau) et la durée moyenne du cycle cellulaire (elle peut être calculée par une expérience d'incorporation de thymidine tritiée suivie d'une autoradiographie : la durée du cycle cellulaire étant égale au temps nécessaire pour obtenir tous les noyaux marqués). Des mesures effectuées chez différents organismes montrent que, d'une façon générale, la phase G1 a la durée la plus variable du cycle cellulaire, et cette variabilité est responsable des différences de durée des cycles cellulaires observés.

La mise en évidence d'un contrôle du cycle cellulaire

Cette mise en évidence s'est faite tout d'abord chez les Levures (*Saccharomyces cerevisiae*, Levure bourgeonnante, et *Schizosaccharomyces pombe*, Levure fissipare). L'étude des mutants présentant des perturbations du cycle cellulaire a permis la découverte de gènes codant pour des protéines impliquées dans le contrôle du cycle cellulaire. Ces protéines étant très conservées au cours de l'évolution, on les a ensuite recherchées dans les cellules de Mammifères et d'autres Vertébrés, ce qui a permis d'identifier les gènes impliqués dans le contrôle du cycle cellulaire chez l'Homme, la Souris, le Xénope, etc.

Les mutations étudiées chez les Levures

S. cerevisiae et *S. pombe* ont des cycles cellulaires similaires. Leur reproduction possible à l'état haploïde a facilité l'étude des mutations. D'autre part, pour maintenir une taille moyenne constante, la durée du cycle cellulaire doit égaler la durée de croissance nécessaire au doublement de la taille de la cellule. Or, la croissance cellulaire dépend beaucoup des nutriments disponibles dans l'environnement ; il existe donc un contrôle du cycle cellulaire en fonction de la taille cellulaire chez les Levures : la cellule ne pourra se diviser que si elle a acquis une taille suffisante. Des mutations dans les gènes responsables de ce contrôle entraînent l'existence de cellules de taille anormale, trop grandes ou trop petites.

Les mutations de deux types de gènes ont été étudiées :

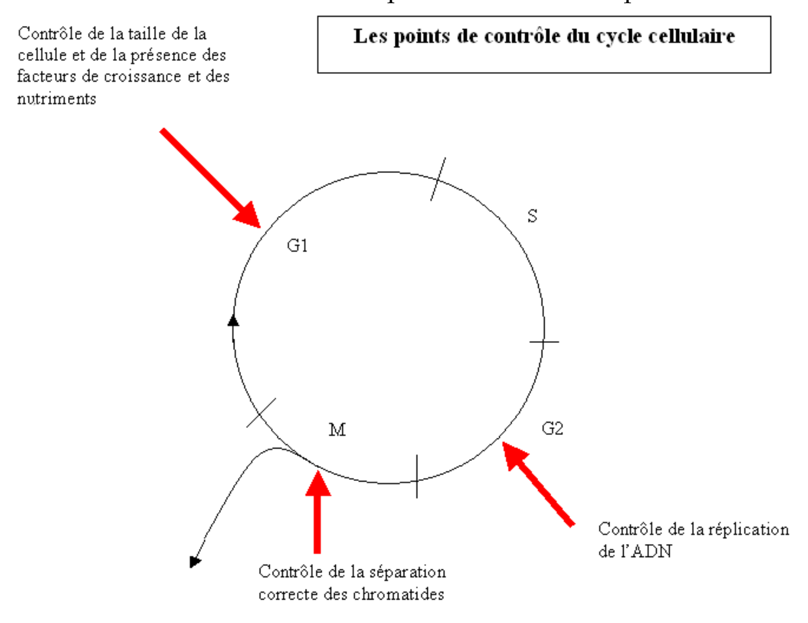
- celles des gènes cdc (Cell division cycle) qui codent pour des protéines kinases qui, en association avec des cyclines spécifiques, déclenchent le passage d'un stade du cycle cellulaire au suivant (par exemple, le passage du stade G2 au stade M pour la protéine kinase codée par le gène cdc2). Les mutations de ces gènes qui ont été le plus utiles pour décrypter le cycle cellulaire sont des mutations conduisant à des Levures dont le phénotype est conditionnel : la protéine « mutante » est fonctionnelle uniquement dans des conditions spécifiques. La plupart des mutants sont thermosensibles : la protéine mutante n'est pas fonctionnelle à température élevée mais permet le déroulement du cycle cellulaire à basse température. Une souche cellulaire de mutant cdc2 thermosensible peut donc pousser uniquement à basse température, condition permissive ;
- celles des gènes qui codent pour des protéines qui agissent sur les protéines codées par les gènes cdc soit en les inhibant, soit en les activant (par phosphorylation ou déphosphorylation). L'expression de ces gènes dépend de facteurs divers, notamment de la taille de la cellule. Par exemple, le gène wee code pour une protéine qui inhibe la protéine kinase codée par le gène cdc2. Tant que cette protéine inhibitrice est produite, le passage du stade G2 au stade M ne peut avoir lieu. Lorsque la taille de la Levure est suffisante, le gène wee cesse de s'exprimer et l'entrée en mitose peut avoir lieu. Un mutant wee déficient ne bloque pas la protéine kinase codée par cdc2 de sorte que la levure entre en mitose de façon prématurée (Levures fissipares de petite taille ; voir le thème de seconde).

C'est chez la Levure fissipare (*S. pombe*) que le gène cdc2 a été découvert. Une mutation de ce gène empêche la mitose : il provoque un arrêt du cycle en empêchant l'entrée en phase G1 ou le passage G2/M (l'activité même du gène cdc2 est régulée par les produits de l'expression d'autres gènes, notamment les gènes wee1, cdc25 et cdc1). Chez la Levure bourgeonnante, le gène cdc28 est équivalent au gène cdc2 de la Levure fissipare (on regroupe maintenant les deux protéines produites par chacun de ces gènes cdc2 et cdc28 sous le terme p34cdc2).

Les points de contrôle du cycle cellulaire

Il existe trois points de contrôle principaux :

- au début de la phase G1 : les facteurs de croissance et les nutriments doivent être présents, la taille de la cellule doit être suffisante, pour que le cycle se poursuive ;
- au niveau du passage S/G2 : les cellules stoppent leur cycle cellulaire si l'ADN est incomplètement répliqué ou endommagé et le reprenne après réparation ;
- au cours de la mitose, à la fin de la métaphase : les cellules vérifient que les chromosomes sont correctement attachés aux fibres du fuseau avant que ne démarre la séparation des chromatides.



(Si des dommages dans l'ADN ou les chromosomes sont constatés et qu'ils ne sont pas réparables, la cellule s'engage sur la voie de l'apoptose ou mort cellulaire programmée.)

Ces points de contrôle agissent sur la poursuite ou non du cycle cellulaire par l'intermédiaire de protéines kinases cyclines dépendantes. Celles-ci résultent de l'association de protéines kinases (cdc ou cdk) avec des cyclines. Il existe de très nombreuses cyclines, et leur taux varie beaucoup au cours du cycle comme leur nom l'indique ; ce taux résulte du bilan entre leur synthèse et leur dégradation. Le taux de protéines kinases cdk (ou cdc) est au contraire relativement constant.

La succession des événements qui aboutit à la division cellulaire est donc en grande partie déterminée par l'expression séquentielle des cyclines ; elles agissent pour initier les phases du cycle cellulaire et aussi pour terminer ces phases.

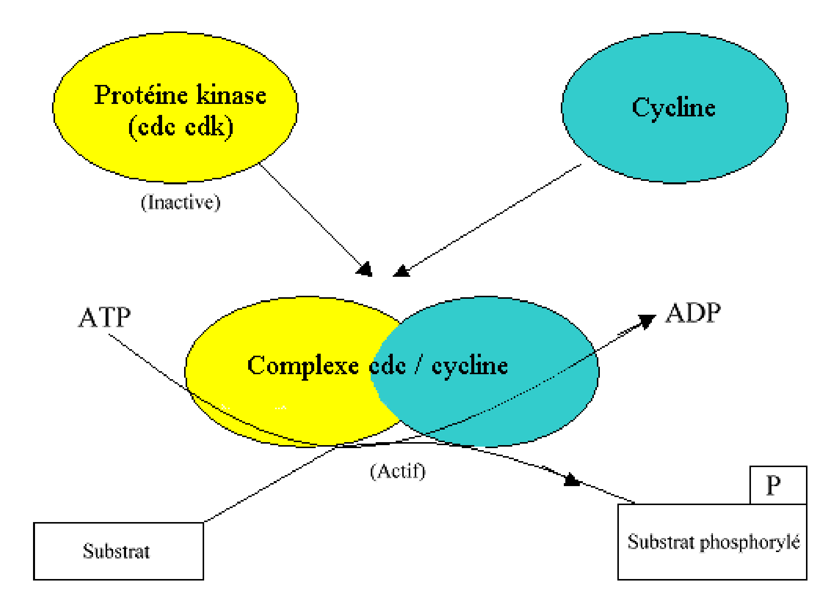
L'activité des protéines kinases cyclines dépendantes est ainsi réglée par le niveau d'expression des cyclines, mais aussi par l'action de protéines les déphosphorylant ou les phosphorylant et par des protéines inhibitrices, les CKI.

Remarques

Les cyclines sont dégradées à la fin de la phase dans laquelle elles interviennent, ce qui évite à la cellule de reprendre la même phase et impose une progression du cycle cellulaire.

Une carence en nutriments réduit la synthèse des cyclines cln par rapport à leur taux de dégradation, ce qui réduit leur concentration qui devient alors inférieure au seuil nécessaire pour entraîner l'activité de la protéine kinase cdc2 (ou p34cdc2). C'est ainsi que l'environnement peut intervenir sur la durée du cycle cellulaire chez la Levure S. *cerevisiae*.

Mode d'action des protéines kinases cycline dépendante (cdc ou cdk)



Les cdk sont donc des enzymes qui catalysent la phosphorylation de protéines cibles jouant un rôle dans les événements du cycle cellulaire : fragmentation de l'enveloppe nucléaire, compaction des chromosomes, etc. De cette phosphorylation résulte un changement de conformation de ces protéines cibles qui entraîne des propriétés nouvelles.

Chez les Levures

Les études réalisées chez la Levure bourgeonnante (*S. cerevisiae*) montrent que selon les phases du cycle, ce n'est pas la même cycline qui sert d'activateur de la protéine kinase cdc2 (ou cdk1) :

- cyclines cln 1, cln 2 cln 3 pour pouvoir démarrer la phase G1 (complexe cdc2/cln1, cdc2/cln2, cdc2/cln3) ;
- cyclines clb 1, clb 2, clb 3, clb 4 pour finir la mitose (complexe cdc2/clb1, cdc2/clb2, cdc2/clb3, cdc2/clb4) ;

Chez la Levure *S. pombe*, elle est impliquée dans le démarrage du cycle en G1 et dans la transition G2-M.

Chez la Drosophile, le Xénope et les Mammifères

Plusieurs associations cycline/protéine kinase contrôlent le cycle cellulaire :

- cdk4/cycline D et cdk2/cycline E pour pouvoir démarrer la phase G1 ;
- cdk2/cycline A pour entrer en phase S ;
- cdk1/cycline B pour entrer en mitose. L'activité du complexe cdk1/cycline B permet la condensation de l'ADN, le désassemblage de l'enveloppe nucléaire et le réarrangement du cytosquelette. En fin de mitose, l'inactivation de ce complexe par d'autres protéines kinases (wee1, myt 1) permet la fin de la mitose.

Le gène cdc2 (ou cdk1)

Le gène cdc2 (*cell division cycle*), chez la Levure, ou cdk1 (*cyclin dependant kinase*), chez les Vertébrés, code pour une protéine kinase cycline dépendante, c'est-à-dire qui a besoin, pour être activée, de se lier à une cycline.

Il est présent chez tous les Eucaryotes et il a été isolé et séquencé chez de nombreux organismes unicellulaires ou pluricellulaires. L'homologie est très importante entre tous ces gènes.

D'autre part, des expériences de transgenèse ont montré que le gène cdc2 d'une espèce, transféré dans une cellule d'une autre espèce, peut y régir le cycle cellulaire : ainsi, des Levures thermosensibles de la souche *Schizosaccharomyces Pombe* dans lesquelles on a transféré le gène cdc2 humain deviennent capables de se diviser à des températures élevées.

Ces observations et expériences sont en faveur d'une origine commune et ancienne des mécanismes de contrôle du cycle cellulaire et mettent en évidence la grande unité du monde vivant.

Les travaux sur le gène CDC2 et sur les autres gènes régulateurs du cycle cellulaire ont valu à P. Nurse, L. Harwell et T. Hunt le prix Nobel de médecine 2001.

Pistes d'exploitation pédagogique des données fournies

Les données moléculaires peuvent être utilisées pour établir des relations de parenté entre les êtres vivants. Ces données s'avèrent bien utiles, notamment quand les autres types de données (anatomiques, morphologiques ou embryologiques) ne sont pas utilisables ; c'est le cas notamment lorsque l'on veut préciser des relations de parenté entre des organismes très différents.

Le gène CDC2 est présent chez tous les organismes eucaryotes, et l'utilisation de données relatives à cet exemple sensibilise à l'idée que les mécanismes fondamentaux de la vie cellulaire sont partagés par tous les organismes eucaryotes.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Parenté entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein du vivant/Gène CDC2** permet d'atteindre :

- **Gène CDC2 chez différentes espèces** qui charge le fichier genescdc2.edi affichant les séquences nucléiques strictement codantes du gène CDC2 chez quelques êtres vivants : Levure, Arabette, Chou, Blé, Maïs, Drosophile, Homme, Xénope, Rat, Souris, Poisson rouge, Grenouille, Poulet, Étoile de mer, Oursin ;

- **Protéine CDC2 chez différentes espèces** qui charge le fichier protCDC2.edi affichant les séquences protéiques du CDC2 correspondant aux séquences nucléiques.

Documents fournis

Dans la banque de documents, le chemin **Parenté entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein du vivant** permet d'atteindre **Gène CDC2** pour charger le fichier cdc2.bmp affichant un texte présentant l'importance du gène CDC2 et précisant son rôle. Ce document apporte un argument supplémentaire à l'unité du vivant en évoquant des expériences de transgenèse réussies.

La présence du gène CDC2 chez tous ces êtres vivants traduit une origine commune. La comparaison deux à deux des séquences nucléiques d'une part et protéiques d'autre part permet d'établir le tableau présenté à la page suivante.

	Arabette	Blé	Chou	Drosophile	Étoile de mer	Grenouille	Homme	Levure	Maïs	Oursin	Poisson rouge	Poulet	Rat	Xénope
Arabette	100													
Blé	82,7	100												
Chou	96,3	82	100											
Drosophile	62	60,9	63	100										
Étoile de mer	63,3	62,7	63,3	69	100									
Grenouille	58,3	57	57,9	62,9	65,2	100								
Homme	65,7	63,6	66	71,7	74,1	78,8	100							
Levure	62,6	62,6	63	60,6	63	58,9	63,6	100						
Maïs	81,6	92,9	81	61,6	63,3	59,5	64,6	63,3	100					
Oursin	60,5	60,1	61,1	68,4	75,1	64,1	72,1	59,8	60,1	100				
Poisson rouge	63,9	63,6	64,6	69,2	74,8	77,2	82,8	61,9	63,9	70,2	100			
Poulet	65,3	62	64,4	69	71,6	77,9	91,4	62	62,4	70	82,8	100		
Rat	65,3	64	65,7	71,7	73,7	79,1	97,6	64	63,6	72,7	83,8	91,6	100	
Xénope	65,2	62,9	64,6	68,9	71,5	87,7	86,8	61,9	63,6	71,2	84,8	87,4	85,4	100

Matrice des identités (en %) entre les protéines cdc2 fournies (obtenue à partir d'un alignement avec discontinuité)

NB : Ces résultats ont été obtenus en choisissant une séquence de référence (la première) et en maintenant toutes les autres dans l'ordre alphabétique des noms. Du fait de l'algorithme de comparaison, ces résultats peuvent être légèrement différents de ceux obtenus par comparaison des séquences deux à deux.

Remarques

- D'un point de vue pratique :

  pour obtenir le degré d'identité entre les séquences comparées, il faut cliquer dans la première case de la ligne traitement, de façon à y faire apparaître la petite flèche rouge. Il faut ensuite cliquer sur l'icône avec un « I » qui donne l'information sur la ligne pointée. Le degré d'identité est exprimé en pourcentage.

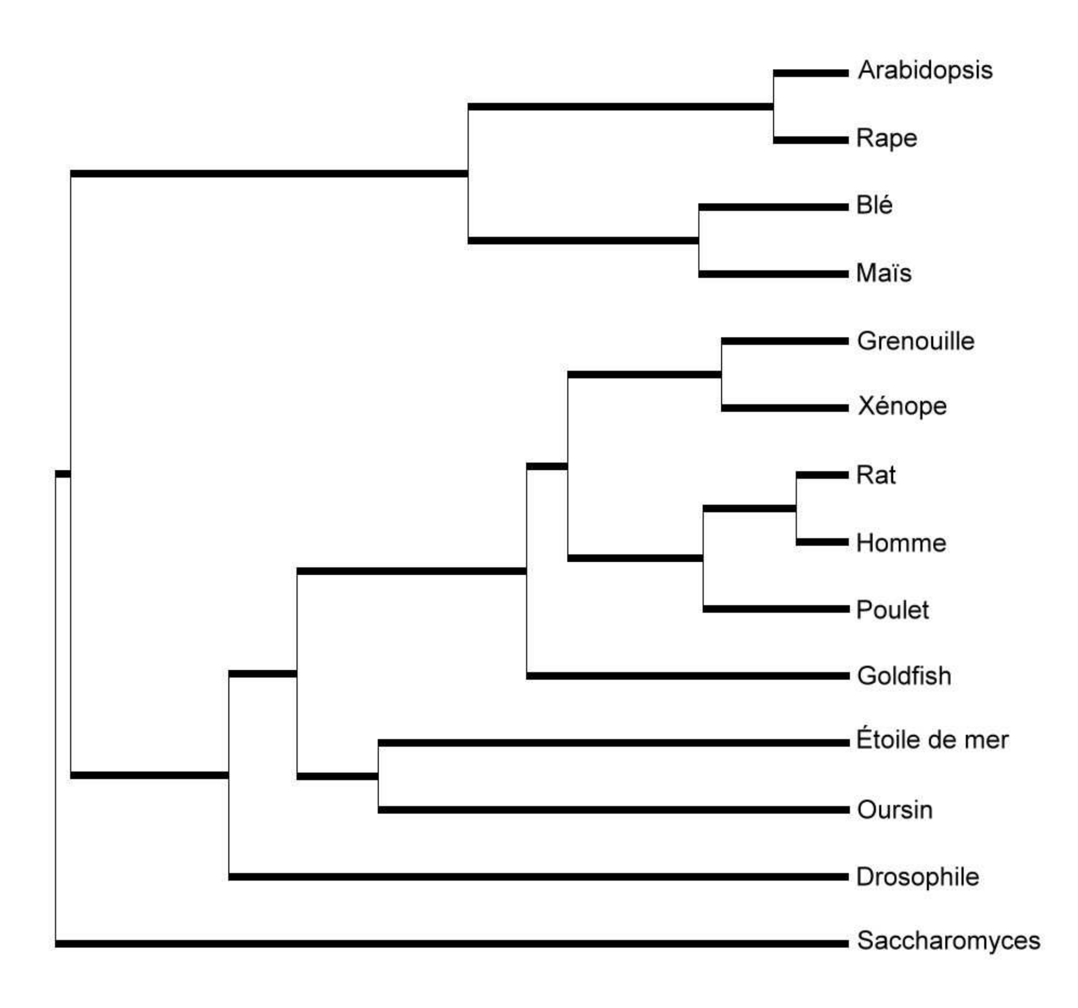
- D'un point de vue conceptuel :

  pour établir une phylogénie à partir de données moléculaires, on commence toujours par relier les éléments qui sont le plus proches, qui présentent le moins de différences, car ce sont eux qui ont l'ancêtre commun le plus récent.

La prise en compte du degré d'identité permet ensuite de préciser les parentés entre ces molécules, donc entre les organismes qui les possèdent. On applique alors une méthode phénétique, basée sur le seul nombre d'identités, sans tenir compte d'une quelconque polarité des états. Le raisonnement est le suivant : les différences constatées sont dues à des mutations ; or ces mutations s'accumulent au cours du temps et on peut donc considérer que plus il y a de différences, et plus de temps s'est écoulé depuis le dernier ancêtre commun,

et inversement ; donc, on peut finalement dire que plus le degré d'identité est important et plus la parenté est grande.

À partir de ces matrices, on peut construire un arbre phylogénétique semblable au suivant obtenu avec le logiciel *Phylogène* (représentation par la méthode UPGMA) :



La construction manuelle d'un tel arbre serait longue et fastidieuse. On peut se limiter à un choix d'espèces en fonction de problèmes à résoudre :

- rechercher qui des Insectes ou des Échinodermes sont les plus proches parents des Vertébrés ;
- établir les relations de parenté entre les quatre plantes à fleurs et voir si elles confortent le rangement de ces plantes dans deux groupes, les Crucifères et les Poacées (Graminées), groupes établis à partir des caractéristiques des Fleurs notamment ;
- rechercher si la classification des Vertébrés à partir de ces données moléculaires est conforme à celle obtenue à partir de données morphologiques et anatomiques.

Relations de parenté au sein du vivant – Gènes homéotiques

Informations scientifiques

Cf. page 45 (classe de seconde)

Pistes d'exploitation pédagogique des données fournies

L'exploitation des données sur les homéoboîtes et les homéodomaines des gènes homéotiques permet d'aborder la notion d'unité du vivant et d'établir des relations de parenté entre certains organismes. Ainsi, des parentés pourront être établies entre des organismes très différents tels un Vertébré, un Arthropode, un Crustacé ou un Ver parasite, alors que les données anatomiques et morphologiques étaient difficilement utilisables.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Parenté entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein du vivant/Gènes homéotiques** pour accéder à :

- **Homéobox (nucléique)** afin de charger les divers fichiers Bx-box.edi affichant les séquences nucléiques des homéoboîtes de quelques gènes homéotiques de Drosophile (Antp, Dfd, Lab, Scr, Pb, Ubx, AbdA, AbdB), d'Homme (Hox B4, B6, B7, B9, C6), de Souris (Hox B4, B6, B7, B9, C6), de Xénope (Hox B4, B7) ;
- **Homéodomaine (protéique)** afin de charger les divers fichiers Bx-domaines.edi affichant les séquences des domaines protéiques de quelques gènes homéotiques de Drosophile (gènes AbdA, AbdB, Antp, Lab, Scr, Dfd, Pb et Ubx), de Souris (Hox B4, B6, C6, B7, B9), d'Homme (Hox B4, B6, C6, B7, B9), de Xénope (Hox B4, B7, B9), de Sacculine (AbdB, Antp) et de Dicyemie (Antp).

La Souris et l'Homme sont des Vertébrés mammifères, le Xénope est un Vertébré amphibien, la Drosophile est un Arthropode Insecte, la Sacculine est un Arthropode crustacé, le Dicyemie est un animal parasite des Calmars et des Pieuvres, très spécialisé, Métazoaire du groupe des Lophotrochozoaires (ce groupe inclut les Annélides, les Brachiopodes, les Mollusques et les Bryozoaires).

Documents fournis

Dans la banque de documents, le chemin **Parenté entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein du vivant/Gènes homéotiques** permet d'atteindre :

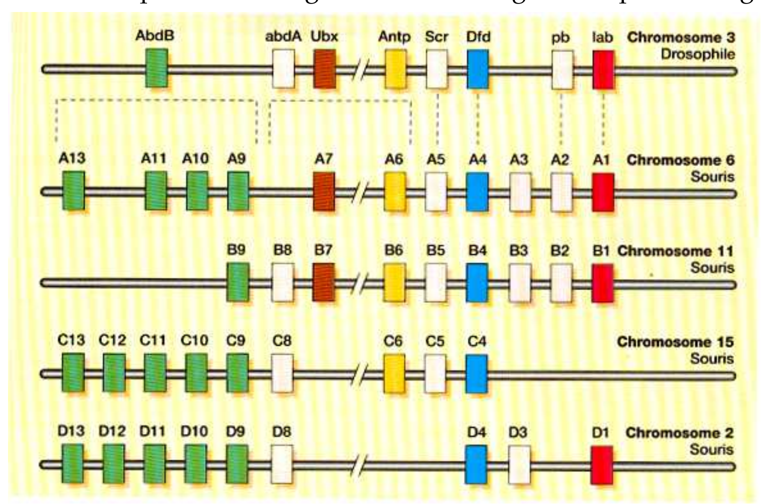
- **Mutants homéotiques de Drosophile** qui charge le fichier phenotypeshomeo.jpg présentant la description des phénotypes de mutants homéotiques permet de comprendre l'action des gènes homéotiques. Ces mutations entraînent la formation d'un organe parfaitement constitué, mais pas au bon endroit. Les gènes homéotiques sont donc des gènes qui vont intervenir au cours du développement embryonnaire pour permettre la mise en place des différents organes. Ils sont responsables de l'édification du plan d'organisation de l'organisme ;
- **Localisation chromosomique** qui charge le fichier chromosomeshomeo.jpg affichant la disposition des gènes sur les chromosomes et permettant de faire plusieurs constats : les gènes sont organisés en complexes (un complexe de gènes disposés en deux groupes appelés parfois complexes antennapedia et bithorax, situés sur le même chromosome chez la Drosophile, et quatre complexes situés sur des chromosomes différents chez la Souris), la disposition des gènes sur les chromosomes correspond à la disposition des régions d'expression de ces gènes dans l'organisme, il y a une correspondance qui a pu être établie entre les gènes de la Drosophile et les gènes de la Souris (ces correspondances sont indiquées par les couleurs) ;
- **Expérience de transgenèse** qui charge le fichier transgenese.bmp : ces expériences de transgenèse réussies sont un argument supplémentaire en faveur de l'homologie des gènes homéotiques de Vertébrés et d'Arthropodes, et donc de l'unité du vivant ;
- **Complexe Homeodomaine/ADN** qui charge le fichier complexeADNHomeo.jpg illustrant le mode d'action des gènes homéotiques. La figure permet de comprendre comment agissent les gènes homéotiques : ils codent pour une protéine dont une partie, appelée homéodomaine, est capable de se fixer sur l'ADN, permettant ainsi la régulation de l'expression de certains gènes.

On se limite ici aux séquences nucléiques des homéoboîtes et des séquences protéiques des homéodomaines car leur similitude est très grande d'un gène homéotique à l'autre, alors que cette homologie est impossible à mettre en évidence si l'on compare les séquences codantes entières de ces gènes (ou les protéines entières). En fait, les gènes homéotiques ne se ressemblent qu'au niveau des homéoboîtes codant pour les homéodomaines protéiques. D'une façon générale, les conclusions obtenues seront les mêmes, que l'on travaille sur la comparaison des séquences protéiques des homéodomaines ou sur la comparaison des séquences nucléiques des homéoboîtes.

Cette similitude de séquences suggère une origine commune, c'est-à-dire une séquence ancestrale commune, donc une origine commune pour les êtres vivants qui les possèdent, c'est-à-dire l'Homme, la Souris, le Xénope, la Drosophile, la Sacculine et le Dicyemie.

Au-delà de l'idée d'origine commune des animaux, cette similitude permet d'imaginer une apparition très précoce des gènes gouvernant la mise en place du plan d'organisation.

On compare les gènes qui sont « équivalents » chez la Drosophile et les Vertébrés. Ces « équivalences » peuvent être repérées sur le document présentant la disposition des gènes homéotiques chez la Drosophile et chez la Souris. Ainsi, on pourra comparer le gène Dfd de la Drosophile avec les gènes hoxA4, B4, C4 et D4 des Vertébrés, le gène Ubx de la Drosophile avec les gènes A7 et B7 le gène Antp avec les gènes A6, B6 et C6.



Comparaison des domaines protéiques

Résultat des comparaisons des domaines protéiques codés par les gènes Dfd de Drosophile et Hox-B4 de trois Vertébrés : on obtient 86,7 % d'identité globale.

Le tableau ci-dessous précise les identités entre les séquences prises deux à deux :

	DP-B4 Homme	DP-B4 Souris	DP-B4 Xénope	DP-Dfd Drosophile
DP-B4 Homme	100 %			
DP-B4 Souris	100 %	100 %		
DP-B4 Xénope	95 %	95 %	100 %	
DP-Dfd Drosophile	88,3 %	88,3 %	88,3 %	100 %

Résultat des comparaisons des domaines protéiques codés par les gènes Ubx de Drosophile et Hox-B7 de trois Vertébrés : on obtient 88,3 % d'identité globale.

Le tableau ci-dessous précise les identités entre les séquences prises deux à deux :

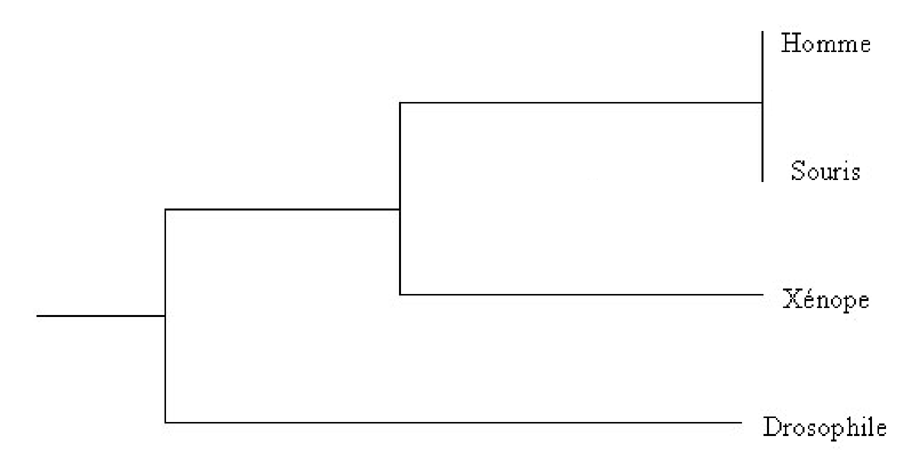
	DP-B7 Homme	DP-B7 Souris	DP-B7 Xénope	DP-Ubx Drosophile
DP-B7 Homme	100 %			
DP-B7 Souris	98,3 %	100 %		
DP-B7 Xénope	96,7 %	96,7 %	100 %	
DP-Ubx Drosophile	90 %	88,3 %	88,3 %	100 %

Résultat des comparaisons des domaines protéiques codés par les gènes AbdB de Drosophile et Hox-B9 de trois Vertébrés : on obtient 86,7 % d'identité globale.

Le tableau ci-dessous précise les identités entre les séquences prises deux à deux :

	DP-B9 Homme	DP-B9 Souris	DP-B9 Xénope	DP-AbdB Drosophile
DP-B9 Homme	100 %			
DP- B9 Souris	100 %	100 %		
DP-B9 Xénope	95 %	95 %	100 %	
DP-AbdB Drosophile	70 %	70 %	70 %	100 %

Si l'on applique la méthode phénétique au traitement de ces données, on peut préciser les parentés entre les êtres vivants concernés : la plus grande parenté étant entre Souris et Homme, puis avec le Xénope, et enfin avec la Drosophile. On peut alors établir la phylogénie suivante :



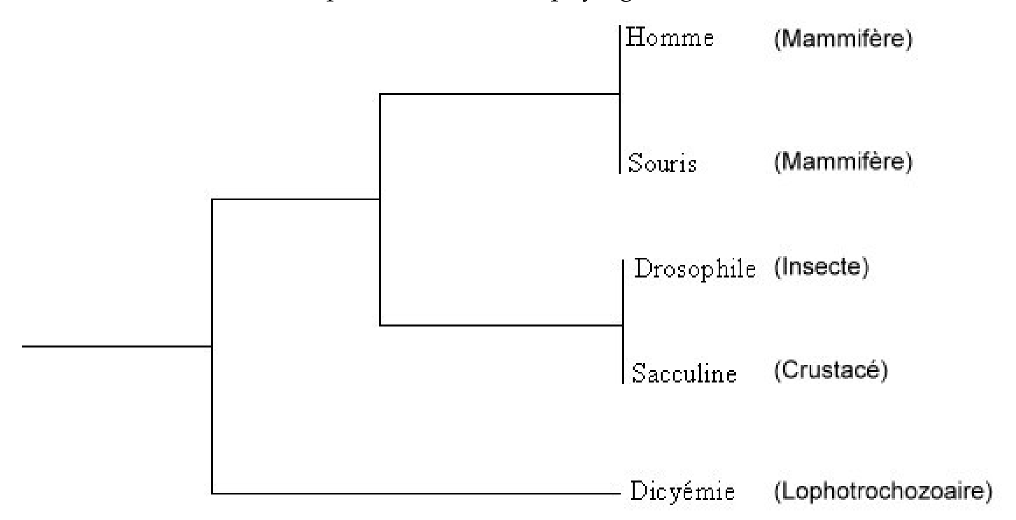
La comparaison des séquences des homéodomaines des gènes Hox B6 de l'Homme et de la Souris avec les gènes Antp de la Drosophile, de la Sacculine et du Dicyemie permet quant à elles d'établir des parentés entre des animaux appartenant à des groupes très éloignés.

Résultat des comparaisons des domaines protéiques codés par les gènes Hox B6 de l'Homme et de la Souris avec ceux codés par les gènes Antp de la Drosophile, de la Sacculine et du Dicyemie : on obtient 71,7 % d'identité globale.

Le tableau ci-dessous précise les identités entre les séquences prises deux à deux :

	DP-B6 Homme	**DP-B6 Souris**	**DP-Antp Drosophile**	**DP-Antp Sacculine**	**DP-Antp Dicyemie**
DP-B6 Homme	100 %				
DP-B6 Souris	100 %	100 %			
DP-Antp Drosophile	93,3 %	93,3 %	100 %		
DP-Antp Sacculine	93,3 %	73,3,3 %	100 %	100 %	
DP-Antp Dicyemie	73,3 %	73,3 %	75 %	75 %	100 %

L'exploitation des valeurs de ce tableau permet d'établir la phylogénie suivante :



Relations de parenté au sein des Vertébrés – Globines alpha et bêta

Informations scientifiques

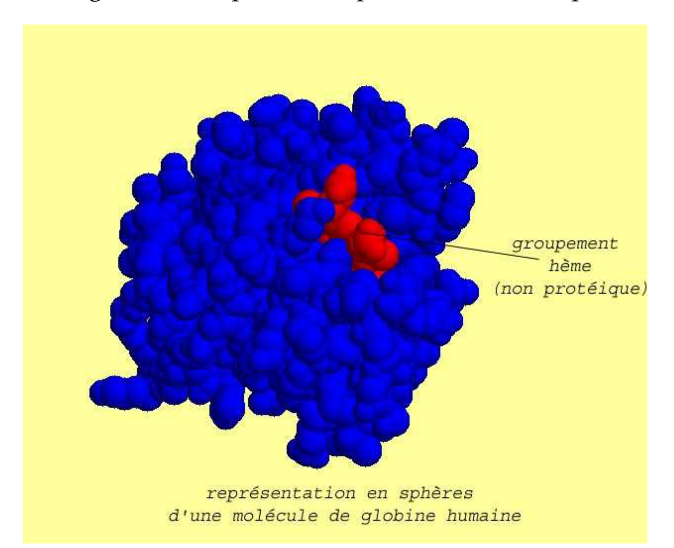
Présentation des globines humaines et des gènes qui les codent

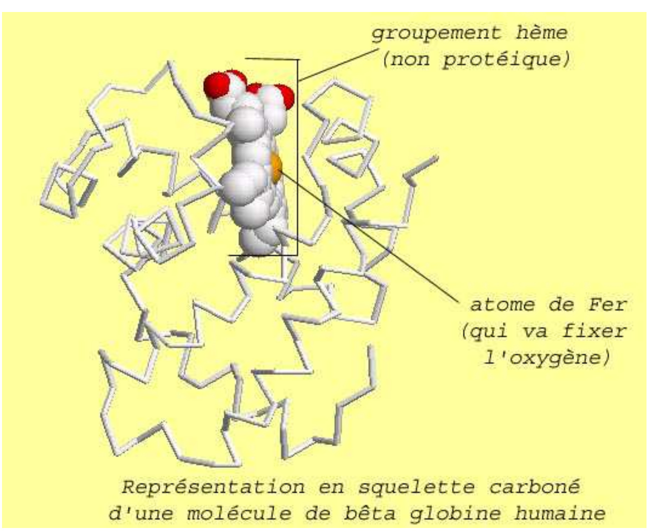
Les globines humaines représentent d'une part les globines qui constituent l'hémoglobine et d'autre part la myoglobine présente dans les cellules musculaires.

Les globines qui constituent les hémoglobines

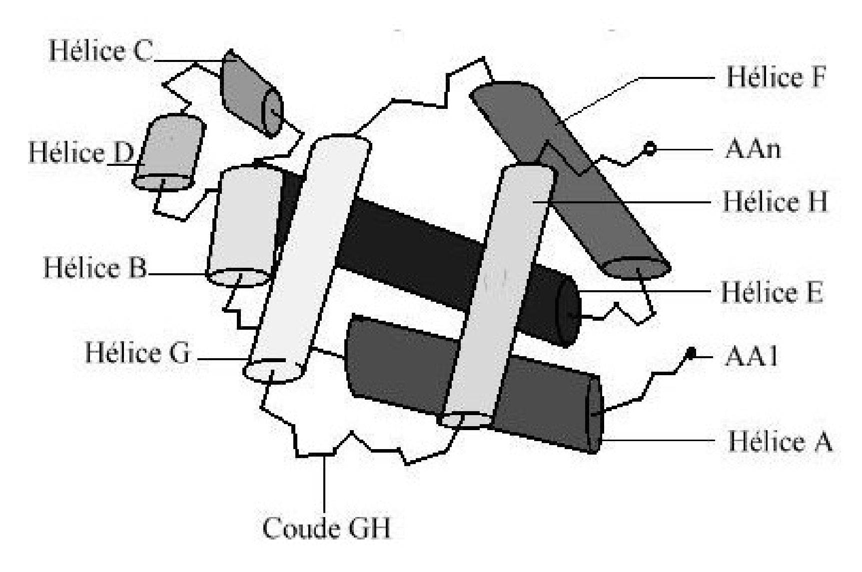
Toutes les hémoglobines humaines sont constituées de quatre chaînes polypeptidiques de globines identiques deux à deux.

Chaque molécule d'hémoglobine est un tétramère formé par l'association de quatre chaînes polypeptidiques identiques deux à deux. Chaque chaîne adopte une conformation spatiale lui donnant une forme globuleuse et ménageant une « poche » superficielle dans laquelle se trouve logé le groupement hème (non protéique).



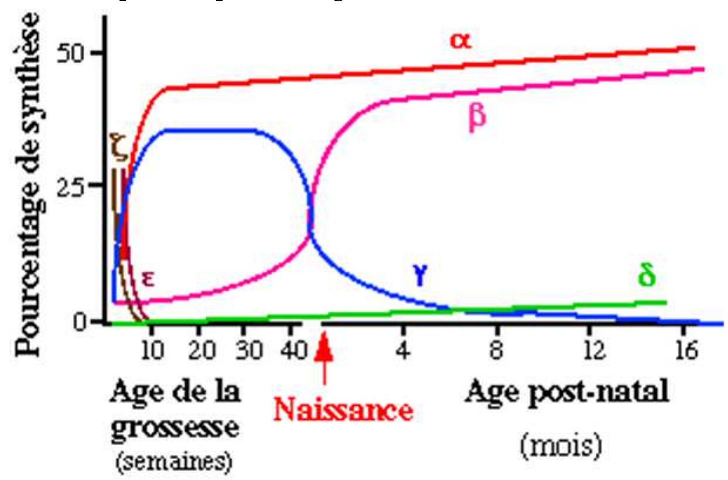


Chaque globine présente sept ou huit secteurs en forme d'hélice droite reliés par des segments comportant parfois des coudes :



Modélisation de la globine bêta humaine

L'être humain, au cours de sa vie, produit plusieurs globines différentes :



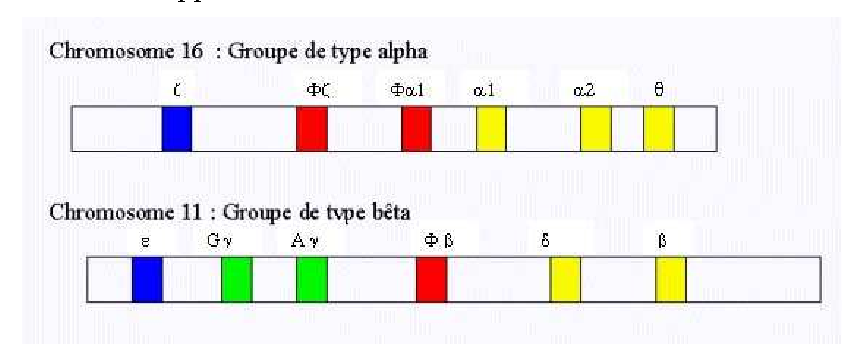
Bien que n'ayant pas la même longueur (141 acides aminés pour les globines alpha 1 et alpha 2, thêta et zêta, 146 acides aminés pour les globines bêta, gamma, delta, epsilon, gamma A et gamma G), toutes ces globines sont fonctionnelles et ont un rôle semblable (fixation et transport du dioxygène grâce au groupement hème).

Avant la naissance, les hématies du fœtus contiennent de l'hémoglobine fœtale constituée de deux chaînes de globine alpha et de deux chaînes de globine gamma. Après la naissance, les hématies contiennent de l'hémoglobine adulte A, très largement majoritaire (97 %), de l'hémoglobine D (2 %) et de l'hémoglobine F (environ 1 %). L'hémoglobine A est constituée de deux chaînes de globine alpha et de deux chaînes de globine bêta. L'hémoglobine D est constituée de deux chaînes de globine alpha et de deux chaînes de globine delta.

Chaque globine est codée par un gène. Les gènes des globines sont répartis sur deux chromosomes différents :

- les gènes de type alpha sont regroupés sur un chromosome (le chromosome numéro 16) ;
- les gènes de type bêta sur un autre (le chromosome numéro 11).

La disposition des gènes de la globine humaine le long des chromosomes correspond à l'ordre dans lequel ils sont exprimés au cours du développement.



Cartes de la famille des gènes codant les chaînes de type alpha et bêta de l'hémoglobine humaine

Les gènes embryonnaires sont représentés en bleu, les gènes foetaux en vert, les gènes adultes en jaune et les pseudogènes en rouge :

- Sur le chromosome 16 on trouve de gauche à droite les gènes **zêta** en bleu, **fz** et **fa**1 en rouge et **alpha1** puis **alpha2** et **thêta** en jaune ;
- Sur le chromosome 11, on trouve de gauche à droite les gènes **epsilon** en bleu, **Gamma G** et **gamma A** en vert, **fb** en rouge, **delta** et **bêta** en jaune.

Les pseudogènes ne codent pas pour des protéines fonctionnelles ; la maturation de l'ARNm ou sa traduction sont bloquées.

Tous les Vertébrés à l'exception des Lamproies et Myxines possèdent au moins deux gènes de globine : le gène bêta et le gène alpha. Les protéines d'alpha globine et de bêta globine constituant les molécules d'hémoglobine ont la même fonction chez tous : la fixation et le transport de dioxygène. Toutes les molécules de bêta globine des Vertébrés sont homologues ; il en est de même des molécules d'alpha globine.

Pistes d'exploitation pédagogique des données fournies

En classe terminale, les parentés entre les Vertébrés peuvent être établies à partir des séquences nucléiques et/ou protéiques.

Les données fournies permettent de préciser les parentés entre quelques Vertébrés appartenant à différents groupes à partir de données moléculaires. Ces activités de construction de relations de parenté à partir de données moléculaires viennent en complément des activités de construction d'arbres phylogénétiques à partir de données anatomiques et/ou morphologiques. L'exploitation des données moléculaires permet :

- de confirmer les relations de parenté entre les Vertébrés établies à partir des autres données ;
- de préciser des relations de parenté lorsqu'il y a discussion (par exemple en ce qui concerne les parentés entre Oiseaux et Reptiles).

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Parenté entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein des Vertébrés/Globines alpha et bêta permet d'atteindre** :

- **Globine alpha Vertébrés** qui charge le fichier globines-alpha-vertebres.edi affichant les séquences protéiques des alpha globines de douze Vertébrés – Homme, Bonobo, Gorille, Chien, Poule, Canard, Alligator, Caïman, Crocodile, Triton, Thon, Raie ;

- **Globine bêta Vertébrés** qui charge le fichier globines-beta-vertebres.edi affichant les séquences protéiques des bêta globines de douze Vertébrés – Homme, Bonobo, Gorille, Chien, Poule, Canard, Alligator, Caïman, Crocodile, Triton, Thon, Raie.

Documents fournis

Dans la banque de documents, le chemin Parenté **entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein des Vertébrés/Globines alpha et bêta** permet d'atteindre **Globines en 3D** qui charge le fichier presentationhb.jpg présentant les molécules de globine alpha et bêta.

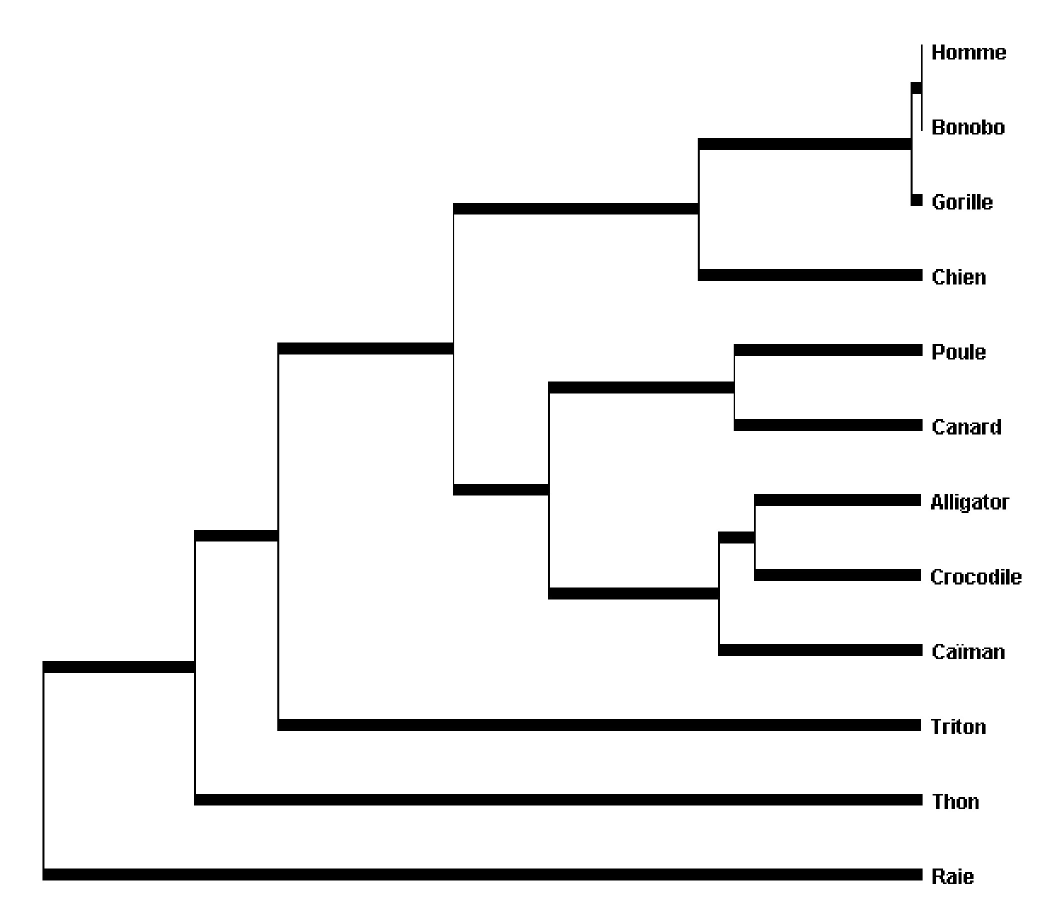
La comparaison des séquences fournies par alignement avec discontinuité permet de remplir une matrice des similitudes. Les résultats et conclusions obtenues à partir des données fournies sont présentés ci-dessous.

À partir de la comparaison des alpha globines de quelques Vertébrés

(fichier globines-alpha-vertebres.edi)

	Homme	Bonobo	Gorille	Chien	Poule	Canard	Alligator	Caïman	Crocodile	Triton	Thon	Raie
Homme	100,0											
Bonobo	100,0	100,0										
Gorille	99,3	99,3	100,0									
Chien	83,7	83,7	84,4	100,0								
Poule	70,2	70,2	69,5	66,7	100,0							
Canard	71,6	71,6	70,9	70,2	87,2	100,0						
Alligator	66,7	66,7	66,0	63,8	71,6	74,5	100,0					
Caïman	67,4	67,4	66,7	62,4	73,0	72,3	85,8	100,0				
Crocodile	68,8	68,8	68,1	66,0	75,9	78,0	87,9	85,1	100,0			
Triton	57,4	57,4	57,4	53,2	56,0	55,3	53,9	54,6	56,0	100,0		
Thon	54,5	54,5	54,5	52,4	49,7	51,0	53,1	51,7	51,7	45,5	100,0	
Raie	45,4	45,4	45,4	45,4	41,1	41,1	44,0	44,0	42,6	37,6	37,6	100,0

Matrice des identités (en pourcentage)



Arbre de parenté que l'on peut obtenir à partir de l'exploitation de cette matrice

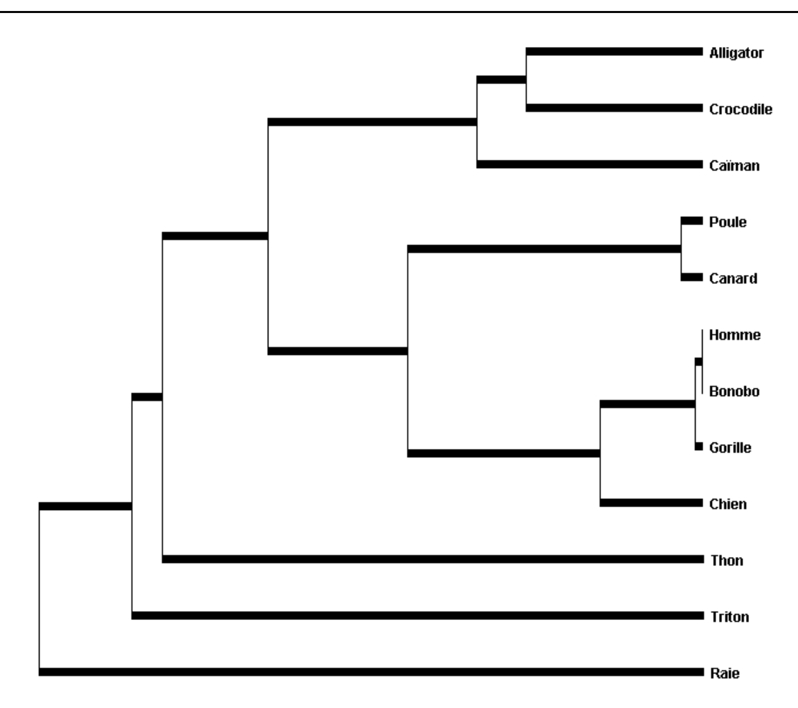
On constate que cet arbre phylogénétique est conforme à celui obtenu à partir de données morphologiques et anatomiques. Aux nœuds successifs de cet arbre, on trouve les ancêtres communs aux Gnathostomes, Ostéichtyens, Tétrapodes, Amniotes, Mammifères et Archosauriens.

À partir de la comparaison des bêta globines de quelques Vertébrés

(fichier globines-beta-vertebres.edi)

	Homme	**Bonobo**	**Gorille**	**Chien**	**Poule**	**Canard**	**Alligator**	**Caïman**	**Crocodile**	**Triton**	**Thon**	**Raie**
Homme	100,0											
Bonobo	100,0	100,0										
Gorille	99,3	99,3	100,0									
Chien	89,7	89,7	90,4	100,0								
Poule	69,2	69,2	68,5	71,2	100,0							
Canard	69,2	69,2	68,5	71,2	97,9	100,0						
Alligator	48,6	48,6	49,3	50,7	61,6	62,3	100,0					
Caïman	49,3	49,3	50,0	52,7	57,5	58,2	79,5	100,0				
Crocodile	54,8	54,8	55,5	55,5	63,7	64,4	80,8	72,6	100,0			
Triton	46,9	46,9	47,6	46,9	46,9	47,6	44,1	42,8	42,8	100,0		
Thon	47,9	47,9	47,3	50,0	54,8	54,8	41,8	41,1	41,1	41,1	100,0	
Raie	38,7	38,7	38,7	40,1	42,3	40,8	33,8	33,1	34,5	30,3	38,0	100,0

Matrice des identités (en pourcentage)



Arbre de parenté que l'on peut obtenir à partir de l'exploitation de cette matrice

L'arbre obtenu à partir des chaînes de globine bêta des mêmes Vertébrés montre certaines des parentés révélées par l'arbre précédent (tous les Mammifères groupés ensemble, Canard et Poule dans le même groupe par exemple). Mais il indique aussi deux différences importantes :

- le Thon est plus étroitement apparenté aux Amniotes que le Triton ;
- les Oiseaux sont plus étroitement apparentés aux Mammifères qu'aux Crocodiliens.

Non seulement l'arbre obtenu à partir des globines bêta est différent de celui des globines alpha, mais de plus, il n'est pas conforme aux données morphologiques et anatomiques. Certaines des informations qu'il traduit sont considérées comme erronées. Cela permet de souligner la limite des informations phylogénétiques extraites des données moléculaires surtout avec la technique d'analyse utilisée au lycée qui est phénétique, c'est-à-dire basée sur le nombre de différences ou de ressemblances entre les molécules. Pour que les informations soient totalement fiables, il faudrait que la vitesse d'évolution des molécules homologues (mutations et fixation des mutations) dans les différentes lignées soit exactement la même. Ce n'est jamais le cas. Lorsque les divergences des lignées sont éloignées dans le temps, il est possible qu'une vitesse plus grande dans une lignée conduise à des résultats inattendus.

COMPLÉMENTS

D'autres fichiers comportant un plus grand nombre de séquences sont fournis. Ils peuvent être utilisés pour créer des sous-ensembles et pour discuter des informations qu'ils fournissent par comparaison à une phylogénie reconnue construite à patir de données morphologiques et anatomiques. Cela peut être utile aussi pour faire travailler des groupes d'élèves sur des exemples différents.

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Parenté entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein des Vertébrés/Globines alpha et bêta** permet d'atteindre :

- **Globines alpha 10 Vertébrés** qui charge le fichier globinesalphareptiles.edi affichant les séquences protéiques des alpha globines de dix Vertébrés, dont sept « Reptiles » (Cobra, Vipère, Iguane, Tortue, Crocodile, Alligator, Caïman), deux Oiseaux (Poule, Canard) et un Batracien (Triton) ;

- **Divers alpha globines** qui charge le fichier alphaglobinetotal.edi affichant les séquences protéiques de l'alpha globine chez de nombreux Vertébrés appartenant à des classes différentes : Mammifères (Homme, Bonobo, Gorille, Panda, Dauphin, Rorqual, Chien, Otarie, Bœuf, Yack, Lama, Souris), Oiseaux (Manchot, Poule, Canard), Reptiles Chéloniens (Tortue géante), Lépidosauriens (Cobra, Vipère, Liophis, Iguane), Crocodiliens (Crocodile, Alligator, Caïman), Poissons Ostéichtyens (Poisson zèbre, Thon), Poissons Chondrichtyens (Raie) ;

- **Divers bêta globines** qui charge le fichier betaglobinetotal.edi affichant les séquences protéiques de la bêta globine chez de nombreux Vertébrés appartenant à des classes différentes : Mammifères (Homme, Chimpanzé, Gorille, Lemur, Rorqual, Dauphin, Chien, Panda), Oiseaux (Poule, Canard, Cormoran), Reptiles Lépidosauriens (Liophis, Varan), Crocodiliens (Crocodile, Alligator, Caïman), Chélonien (Tortue carette), Amphibiens (Triton, Crapaud, Xénope), Poissons Ostéichtyens (Poisson rouge, Thon), Poissons Chondrichtyens (Raie).

Relations de parenté au sein des Primates – Gène de l'opsine S

Informations scientifiques

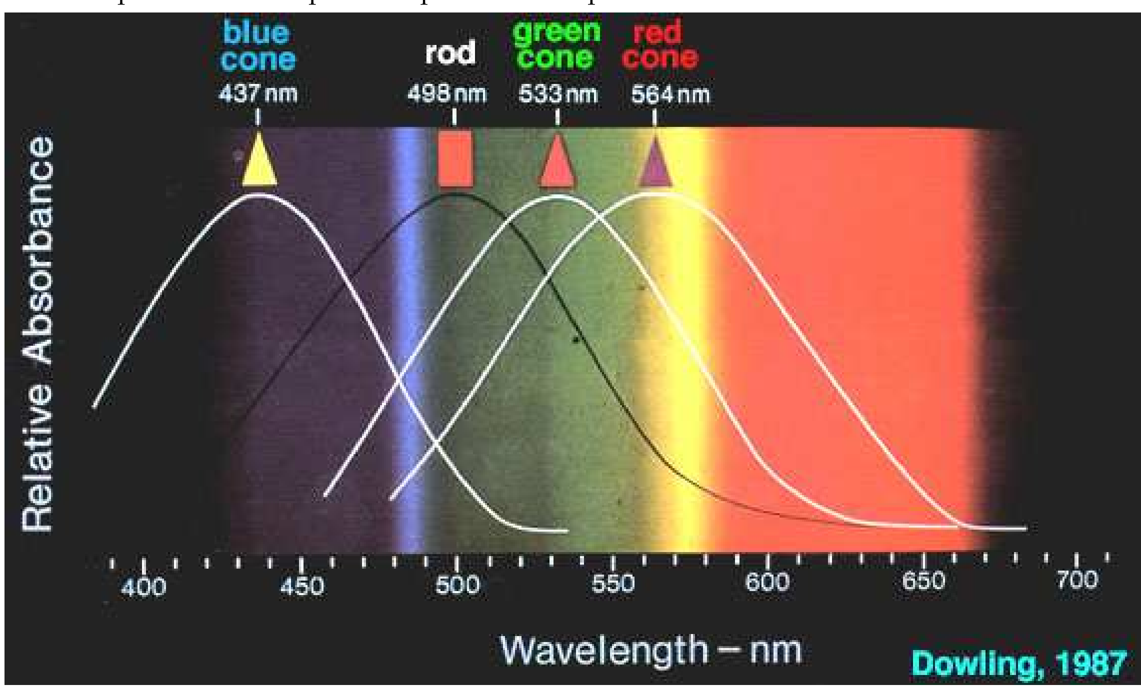
La rétine humaine est un tissu nerveux qui renferme notamment des cellules photoréceptrices de deux types : les cônes et les bâtonnets. Les cônes sont responsables de la vision des couleurs ; les bâtonnets ne permettent pas la vision des couleurs, mais sont très sensibles à la lumière et permettent une vision en très faible éclairement.

La vision des couleurs est liée à l'existence dans la rétine de trois types cellulaires (les cônes) qui sont des photorécepteurs synthétisant trois pigments différents, des opsines, qui présentent des niveaux d'absorption différents dans le rouge (opsine rouge ou L), le vert (opsine verte ou M) ou le bleu (opsine bleue ou S). Avec un seul des pigments, on ne voit pas les couleurs ; la présence des trois pigments permet une vision trichromatique.

Les pigments des cellules photoréceptrices

Toutes ces cellules contiennent un pigment, de nature essentiellement protéique, excitable par des photons. Ce pigment varie selon le type de cellules photoréceptrices :

- les bâtonnets synthétisent un pigment appelé **rhodopsine** (constituée d'une opsine et d'un groupement prosthétique, le 11-cis-rétinal lié à la lysine 216 de la protéine, dont le précurseur est le all-trans-rétinol ou vitamine A. Un déficit alimentaire en vitamine A conduit ainsi à l'héméralopie ou cécité nocturne). La rhodopsine présente une bande d'absorption large dans la région visible du spectre avec un pic à 500 nm.
- les cônes synthétisent un pigment appelé opsine ; il existe trois types de cônes en fonction du pigment qu'ils synthétisent : les cônes à opsine S (Short ou bleue), M (Médium ou verte) et L (Long ou rouge), ces différentes opsines différant par leur spectre d'absorption.



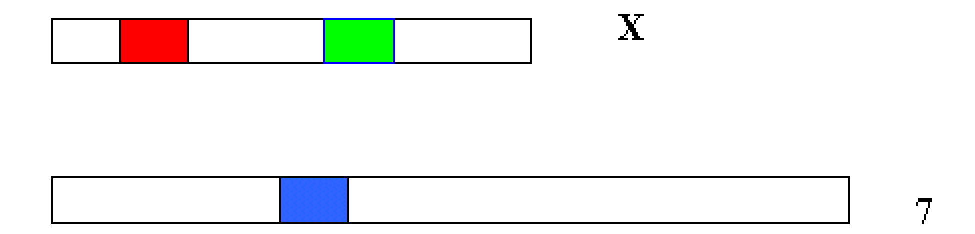
Les spectres d'absorption des trois opsines photoréceptrices ont été mesurés en éclairant des cônes avec un faisceau de lumière de 1 mm seulement. Les réponses des différents cônes montrent trois groupes : certaines sont excitées au maximum par la lumière bleue, d'autres sont excitées au maximum par la lumière verte et les dernières par la lumière orangée-rouge.

Les cônes S contiennent une opsine ayant un maximum d'absorption dans les courtes longueurs d'onde du visible (« opsine S ou bleue »), les cônes M contiennent une opsine ayant un maximum d'absorption dans les longueurs d'onde moyennes du visible (« opsine M ou verte »), les cônes L contiennent une opsine ayant un maximum d'absorption dans les grandes longueurs d'onde du visible (« opsine L ou rouge »). Bien entendu, tous les photorécepteurs possèdent les mêmes gènes codant pour les opsines, mais un seul de ces gènes s'exprime dans un photorécepteur donné.

Les protéines photoréceptrices des cellules en cône sont, comme l'opsine de la rhodopsine, des récepteurs à sept hélices, et elles contiennent elles aussi le 11-cis-rétinal.

Localisation chromosomique des gènes codant pour les opsines humaines :

Les gènes codant pour les opsines M ou verte et L ou rouge sont situés sur le chromosome X. Le gène codant pour l'opsine S ou bleue est porté par le chromosome 7.



Les opsines chez les Primates

Chez les Primates, la vision trichromatique n'est possible que dans certains groupes :

- les Singes de l'Ancien Monde (Afrique, Asie et Europe), ou catarhiniens, possèdent, comme l'Homme :
    - le gène de l'opsine S (bleue) porté par un autosome, codant pour une opsine présentant un pic d'absorption à 435 nm ;
    - le gène de l'opsine M (verte) porté par le chromosome X, codant pour une opsine présentant un pic d'absorption à 535 nm ;
    - le gène de l'opsine L (rouge) porté par le chromosome X, codant pour une opsine présentant un pic d'absorption à 565 nm.

  Ces Singes sont donc tous trichromates ;
- les Singes du Nouveau Monde (Amérique), ou platyrhiniens, possèdent :
    - le gène de l'opsine S (bleue), situé sur un autosome et codant pour une opsine ayant un pic d'absorption à 430 nm environ ;
    - un seul autre gène codant pour une opsine S/M, porté par le chromosome X, codant pour une opsine ayant un pic d'absorption de 530 à 560 nm. Or il existe trois allèles différents possibles pour cette opsine, codant pour des pigments dont les pics d'absorption se situent vers 536 nm, 556 nm, 563 nm.

  Chez ces Singes, comme le Saïmiri par exemple (Singe écureuil), on a ainsi plusieurs phénotypes possibles pour la vision des couleurs (Singes dichromates ou Singes trichromates), notamment en fonction du sexe :
    - les mâles sont tous dichromates : ils possèdent le gène de l'opsine S, et un seul gène pour une opsine M/L car ils n'ont qu'un seul chromosome X ;
    - les femelles, qui possèdent deux chromosomes X, peuvent être dichromates ou trichromates : elles ont toutes un gène de l'opsine S, et possèdent un seul allèle du gène de l'opsine M/L si elles sont homozygotes, mais en possèdent deux si elles sont hétérozygotes. Si les deux allèles qu'elles possèdent sont différents, elles produiront donc trois opsines différentes, ce qui permettra une vision trichromatique.

  On a d'autre part récemment découvert que les Singes hurleurs (*Alouata seniculus*) étaient trichromates, aussi bien les mâles que les femelles. Ils synthétisent trois pigments visuels présentant des pics d'absorption à 430 nm, 532 nm et 562 nm. C'est actuellement le seul Singe du Nouveau Monde pour lequel on ait découvert cette trichromatie.
- les Lémuriens : ceux qui vivent la nuit ont une vision en noir et blanc ; ceux qui vivent le jour ont tous une vision dichromatique. Ils possèdent le gène de l'opsine S (bleue) sur un autosome, et le gène d'une opsine M/L sur le chromosome X.

Pistes d'exploitation pédagogique des données fournies

Il est possible d'établir des relations de parenté entre les Primates à l'aide des données moléculaires concernant les opsines, et notamment l'opsine S (ou opsine bleue). Tous les Primates possèdent ce gène, porté par un autosome, et il code pour une opsine ayant toujours un pic d'absorption voisin de 430 nm.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Parenté entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein des Primates/Gène de l'opsine S (Shortwave - Bleu)** permet d'atteindre **Opsine S bleu Primates** qui charge le fichier OpsPrimates.edi affichant les séquences protéiques des opsines bleues (ou opsines S) de quelques Primates appartenant à des groupes différents :

- platyrhiniens : Cebus (Capucin), Saïmiri (Singe écureuil), Alouate (Singe hurleur) ;
- catarhiniens : Homme, Macaque, Bonobo, Chimpanzé, Gorille.

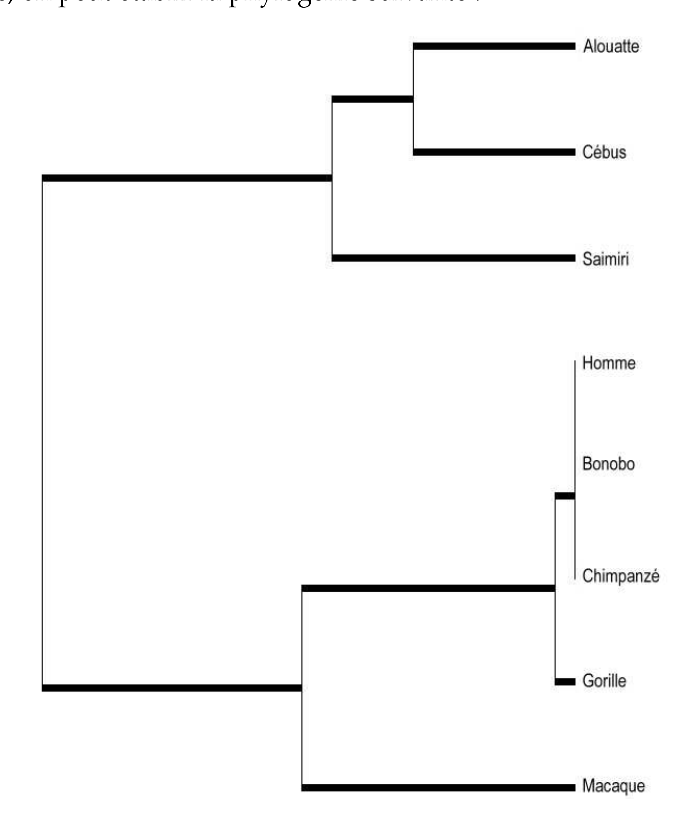
Documents fournis

Dans la banque de documents, le chemin **Parenté entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein des Primates/Gène de l'opsine S (Shortwave – Bleu)** permet d'atteindre **Opsine S** pour afficher le fichier Opsine S.bmp présentant la nature et le rôle de l'opsine S ainsi que la localisation du gène qui la code.

Les informations tirées de la comparaison des séquences par alignement avec discontinuités permettent de construire la matrice des identités suivante (les valeurs sont en pourcentage) :

	Homme	**Bonobo**	**Chimpanzé**	**Gorille**	**Macaque**	**Alouate**	**Saïmiri**	**Cebus**
Homme	100							
Bonobo	100	100						
Chimpanzé	100	100	100					
Gorille	99,7	99,7	99,7	100				
Macaque	96	96	96	95,7	100			
Alouate	92,3	92,3	92,3	92	92,3	100		
Saïmiri	91,7	91,7	91,7	91,4	92,2	97,1	100	
Cebus	92,5	92,5	92,5	92,2	92,5	97,7	96	100

À partir de cette matrice, on peut établir la phylogénie suivante :



Relations de parenté au sein des Hominidés - *ADN mitochondrial*

D'autres informations sur ce thème peuvent être trouvées sur le site *Vie à Jussieu* dont l'adresse est : http://www.snv.jussieu.fr/vie/documents/adnancient/adnmt.htm

Informations scientifiques

ADN fossile et histoire des Homininés

Un arbre phylogénétique des Homininés basé sur l'utilisation de caractères morphologiques et anatomiques peut être construit avec le logiciel *Phylogène*.

Les données moléculaires commencent à être utilisées pour aborder certains problèmes de l'histoire évolutive des Homininés. Sous ce terme, on désigne ici les espèces appartenant à la lignée humaine qui partagent avec *Homo Sapiens* un ancêtre commun qui n'est pas celui des Chimpanzés. L'espèce humaine est la seule espèce d'Homininés qui existe aujourd'hui. Pour exploiter les données moléculaires relatives à d'autres espèces, il faut se procurer de l'ADN de ces espèces, donc recourir à de l'ADN fossile. La première étude date de 1997 et elle est relative à l'Homme de Neandertal.

Après la mort d'un organisme, l'ADN se dégrade rapidement, surtout si la température de la région est assez élevée. L'ADN que l'on peut extraire d'un fossile se trouve sous forme de courts fragments de quelques dizaines à une ou deux centaines de nucléotides. Au-delà de 100 000 ans, il est illusoire de vouloir récupérer de l'ADN utilisable pour déchiffrer une séquence. Cela réduit l'utilisation de l'ADN fossile aux périodes les plus récentes de l'histoire de la lignée humaine, donc aux « espèces » qui sont contemporaines des premiers Sapiens. Pour l'instant, on n'a utilisé que de l'ADN de neandertaliens. *A priori*, il pourrait être possible d'exploiter l'ADN de l'Homme de Flores qui a vécu jusqu'à -12 000 ans, bien que les conditions climatiques de l'Indonésie ne soient pas favorables à la préservation de l'ADN. Si cela s'avérait réalisable, les informations obtenues pourraient éclairer la place controversée de cet Homininé dans l'histoire de la lignée humaine.

Intérêt de l'ADN mitochondrial

Les études sur l'ADN fossile des Homininés ont porté sur l'ADN mitochondrial et non sur l'ADN nucléaire. Cela s'explique pour diverses raisons :

- presque tout l'ADN disparaît au cours de la fossilisation. Il n'en reste que des courts fragments. Le premier travail consiste à amplifier par la technique de la PCR (polymerase chain reaction) un ou plusieurs fragments qui ont été préservés. Il faut choisir d'avance ceux que l'on recherche dans l'extrait fossile. Il y a généralement plusieurs centaines de mitochondries par cellule et chacune contient une molécule circulaire d'ADN de plus de 16 000 paires de bases. Les séquences d'ADN mitochondrial sont donc au nombre de plusieurs centaines dans une cellule alors que celles de l'ADN nucléaire sont au nombre de deux sauf pour les séquences répétées. Il est donc **plus probable de trouver dans l'ADN fossile des fragments d'ADN mitochondrial** susceptibles d'être amplifiés que des fragments d'ADN nucléaire ;
- un autre point important est que l'ADN mitochondrial évolue plus rapidement que l'ADN nucléaire : les mutations s'y fixent plus rapidement. Pour des périodes relativement courtes imposées par la préservation de l'ADN fossile, l'ADN mitochondrial est donc **plus informatif** que l'ADN nucléaire. En outre, le génome mitochondrial ne se transmet que de la mère à l'enfant (les mitochondries du spermatozoïde ne contribuent pas au stock initial de mitochondries du zygote). Il n'y a donc pas de recombinaison. Les différences entre molécules d'ADN mitochondrial résultent uniquement de mutations ;
- le séquençage de l'ADN mitochondrial d'un très grand nombre d'Hommes actuels a révélé que deux régions appartenant à la zone de contrôle de la réplication de cet ADN présentent une grande variabilité : ce sont les régions hypervariables 1 et 2. Ce sont des fragments de ces régions hypervariables qui ont été recherchés chez les neandertaliens.

Objectifs de l'étude de l'ADN mitochondrial fossile

Les recherches sur l'ADN mitochondrial des neandertaliens ont pour principal objectif de déterminer si ces Homininés ont pu contribuer au pool génétique des populations de *Homo sapiens*.

Cette question s'insère dans le problème plus général de l'origine des Hommes modernes. Au cours des années 1980, deux hypothèses extrêmes se sont opposées. Selon l'hypothèse multirégionale, dite encore du candélabre, des populations d'*Homo erectus* d'origine africaine se sont répandues à travers le monde et ont graduellement et indépendamment évolué pour donner d'abord des *Homo Sapiens* archaïques, puis des Hommes modernes. Selon cette hypothèse les Hommes de Neandertal seraient issus *d'Homo Sapiens* archaïques européens (les préneandertaliens) et

seraient à l'origine des *Homo sapiens* européens et du Proche-Orient. Selon la seconde hypothèse, dite de l'arche de Noé, tous les *Homos sapiens* actuels auraient pour origine une population initiale du proche orient ou d'Éthiopie qui aurait migré ensuite dans tout le monde. Cette population résulterait d'un processus de spéciation, il y a 150 000 ans environ. Elle se serait ensuite répandue dans toutes les régions du monde sans se métisser avec les populations d'*Homo sapiens* archaïques déjà présentes. Ces *Homo sapiens sapiens* auraient contribué par des mécanismes mal élucidés à la disparition des *Homo sapiens* archaïques, des neandertaliens en Europe.

Aujourd'hui, il y a un accord assez général pour dire que la première hypothèse, celle du candélabre ne s'applique pas à l'Europe. En effet, on ne connaît pas de fossiles d'*Homo* permettant d'asseoir l'hypothèse d'une transformation graduelle de l'Homme de Neandertal en *Homo sapiens*. Toutefois, au Proche-Orient, les populations de neandertaliens ont cohabité pendant plus de 50 000 ans avec celles des premier *Homo sapiens* ; en Europe, la cohabitation a duré plus de 10 000 ans (de -40 000 ans à -29 000 ans environ, date de disparition des derniers neandertaliens). Au cours de cette période de cohabitation, on ne peut exclure des possibilités de métissage entre les Hommes de Cro-Magnon et les Hommes de Neandertal. Si tel est le cas, l'Homme de Neandertal appartiendrait à la même espèce que l'Homme actuel et devrait s'appeler *Homo sapiens neandertalensis*. En revanche, s'il n'y a pas eu hybridation, l'Homme de Neandertal serait une espèce différente de celle de *Homo sapiens* et mériterait le nom de *Homo neandertalensis*. Dans ce cas, les neandertaliens représenteraient un cul-de-sac évolutif puisque disparus sans laisser de descendance.

C'est à la résolution de cette question, métissage ou non entre neandertaliens et premiers Hommes modernes, que visent les études sur l'ADN mitochondrial.

Apports des séquences d'ADN mitochondrial

C'est en 1997 qu'a été déterminée la première séquence d'ADN mitochondrial d'un neandertalien. Il s'agit de la séquence, longue de 379 paires de nucléotides, de la région hypervariable 1 de la zone où se trouve l'origine de la réplication de l'ADN. Elle a été établie à partir d'un extrait d'ADN du spécimen type de l'Homme de Neandertal, celui de Feldhofer (vallée de Neander, Allemagne) découvert en 1856. Elle a été comparée à 994 séquences homologues d'Hommes actuels appartenant à différentes régions du monde dont celles de 510 Européens. Le nombre moyen de différences entre les séquences des Hommes actuels est de huit. Le nombre moyen de différences entre la séquence de l'Homme de Feldhofer et celles des Hommes modernes est de 27, soit plus de trois fois plus importante. La plus petite différence entre la séquence neandertalienne et celle d'un Homme actuel est de 22 et la plus grande de 36. La séquence neandertalienne se place donc nettement au-delà de la variabilité constatée entre Hommes actuels. Le nombre de différences entre la séquence neandertalienne et celles des Asiatiques et des Africains est similaire. En particulier, aucune séquence d'Européen ne s'est avérée plus proche de celle de l'Homme de Feldhofer. Ces données n'apportent aucun argument en faveur de l'hypothèse suivant laquelle les neandertaliens auraient contribué au pool génétique des Européens par suite de métissage avec les Hommes de Cro-Magnon, en réalité le métissage entre femmes neandertaliennes et Hommes de Cro-Magnon, puisque l'ADN mitochondrial se transmet uniquement par les femmes.

Trois autres séquences de la région hypervariable 1 ont par la suite été déterminées à partir de fossiles neandertaliens d'âges différents et appartenant à des régions de l'Europe très éloignées les unes des autres : Allemagne, Caucase, Croatie. Elles ont entièrement confirmé les indications fournies par la première séquence. De plus, quand on établit un arbre phylogénétique à partir de ces séquences et de celles des Hommes actuels, on constate que les séquences des neandertaliens se groupent entre elles ; les neandertaliens paraissent avoir un ancêtre commun qui n'est pas celui des Hommes modernes. Le nombre moyen de différences entre ces séquences neandertaliennes appartenant à des individus ayant vécu dans des régions européennes éloignées et à des moments différents (-40 000 à –29 000 ans) est du même ordre que celles constatées entre les séquences des Hommes actuels.

Si tout cela ne plaide pas en faveur du métissage entre neandertaliens et Hommes de Cro-Magnon, on ne peut pas écarter l'hypothèse qu'il a eu lieu mais que l'on n'en retrouve pas de signe dans les populations européennes actuelles. En effet, les descendants de ces métissages ont pu appartenir à des populations qui n'ont pas survécu au cours du paroxysme de la dernière période glaciaire. On pouvait donc penser que l'on avait plus de chance de repérer ce métissage dans l'ADN fossile des Hommes de Cro-Magnon qui ont cohabité avec les neandertaliens.

En mai 2003, l'équipe de Caramelli a publié la séquence de la région hypervariable 1 de deux Hommes de Cro-Magnon ayant vécu en Italie, il y a 24 000 ans. Leurs séquences sont comparables à celles des Hommes actuels ; dans un arbre phylogénétique, ces Hommes de Cro-Magnon se retrouvent parmi les Hommes actuels et non avec les neandertaliens.

En 2004, l'équipe de David Serre et d'André Langaney a publié les résultats d'une étude sur l'ADN fossile de neandertaliens (La Chapelle-aux-Saints, Engis) et d'Hommes de Cro-Magnon (cinq spécimens, notamment des sites

de La Madeleine et de Cro-Magnon). Disposant des informations sur les séquences de neandertaliens précédentes, ils ont essayé d'obtenir sur les différents échantillons d'ADN fossile l'amplification de fragments de la région hypervariable 1 par la technique PCR en utilisant des amorces typiques des séquences neandertaliennes. Ils ont obtenu des produits d'amplification pour tous les extraits de neandertaliens mais aucun à partir des extraits de Cro-Magnon.

Ainsi, jusqu'ici, aucune trace neandertalienne n'a été mise en évidence chez les Hommes de Cro-Magnon. Il semble que l'Homme de Neandertal n'a légué aucun tribut génétique à l'Homme moderne.

Pistes d'exploitation pédagogique des données fournies

Séquences et documents

Fichiers des séquences

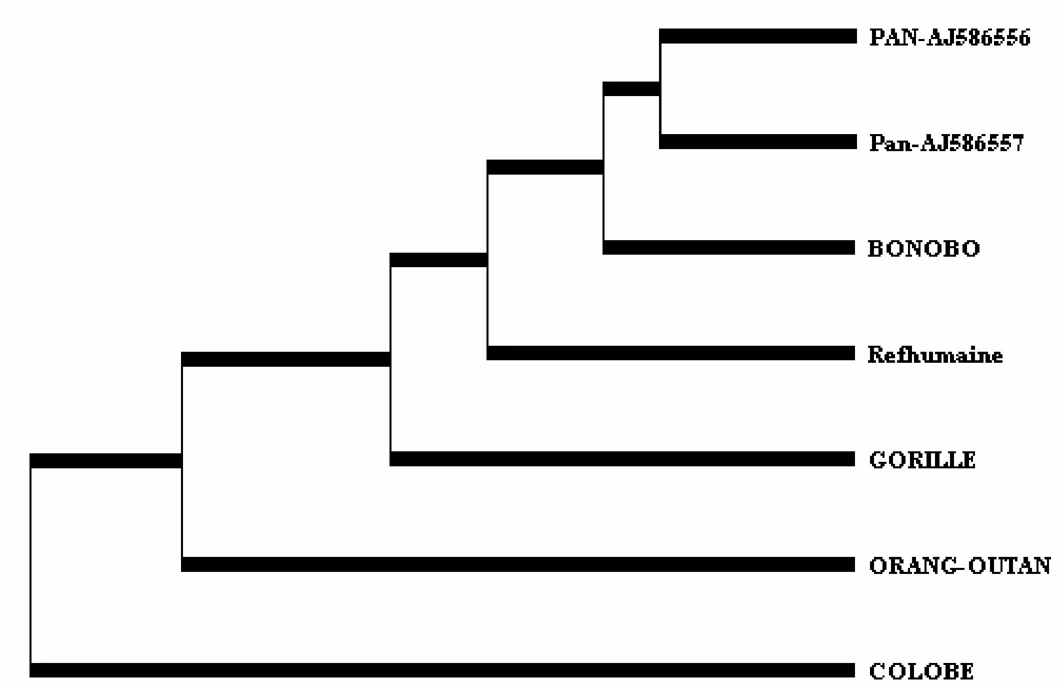
Dans la banque de thèmes d'étude, le chemin **Parenté entre êtres vivants actuels et fossiles – Phylogenèse, évolution/Relations de parenté au sein des Hominidés/ADN mitochondrial** permet d'atteindre :

- **Primates** qui charge le fichier Primates.edi affichant la séquence consensus humaine de référence et de divers Primates ;
- **Actuels et fossiles** qui charge le fichier 7mod+3neand+2cromagnon-2pan.edi affichant les séquences de sept Hommes actuels appartenant à divers continents, trois neandertaliens (deux de Feldhofer, l'autre de Croatie), deux Hommes de Cro-Magnon (ceux d'Italie) et deux Chimpanzés ;
- **ADN mitochondrial partiel** qui charge le fichier ADN-Mito-partiel.edi portant sur une portion de cette séquence de 345 nucléotides, et qui a été élaboré à partir des séquences d'un fragment d'amplification de neandertaliens obtenus par l'équipe de David Serre et d'André Langaney, les affichant pour l'Homme moderne, Cro-Magnon, neandertalien.

Les séquences fournies sont celles de la région hypervariable 1 de l'ADN mitochondrial de divers Primates et d'Hominінés. Afin de pouvoir faire des comparaisons, on a sélectionné des séquences homologues de 345 nucléotides à partir des séquences fournies par la banque internationale sur la région hypervariable 1.

Bien entendu, l'exploitation de ces fichiers nécessite un minimum d'informations préalables sur l'ADN mitochondrial et l'ADN fossile. L'exploitation des données fournies est de même nature que celle des séquences utilisées précédemment pour établir des phylogénies.

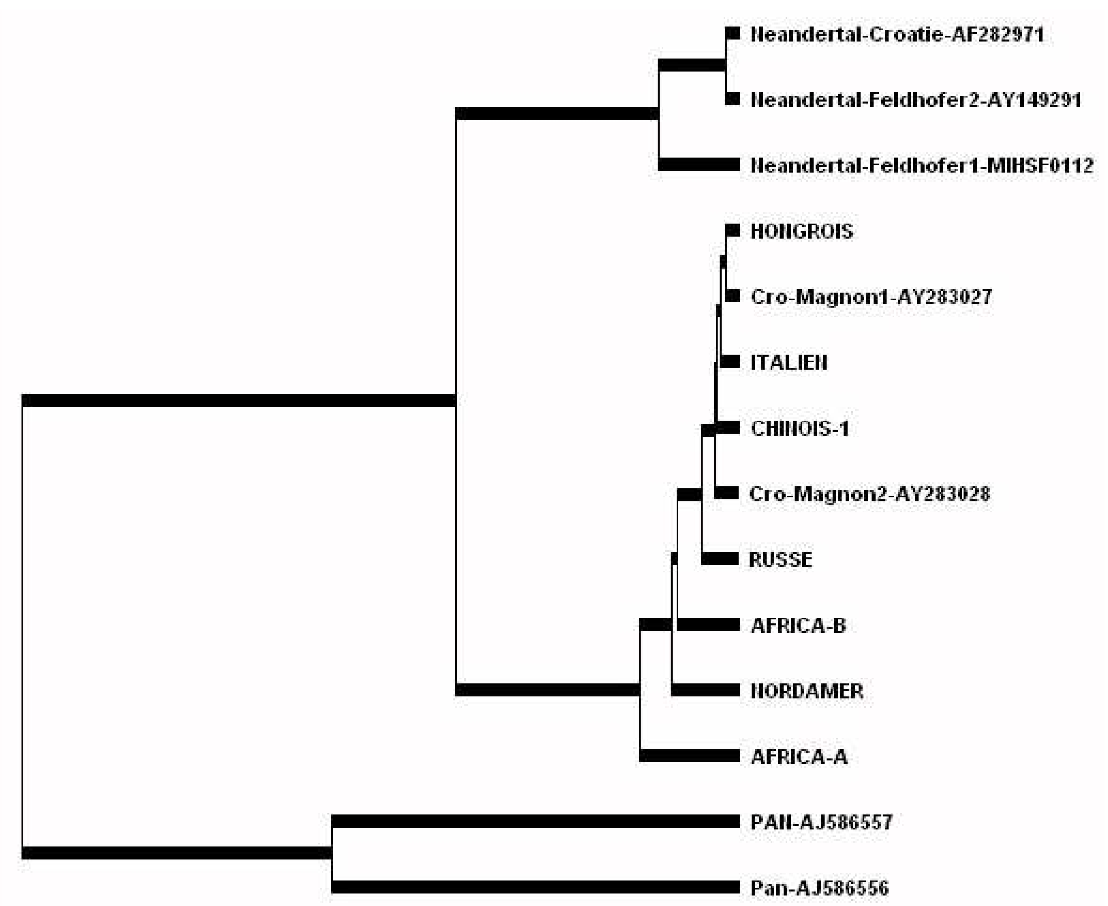
Le fichier sur les Primates permet de discuter la pertinence d'utiliser l'ADN mitochondrial pour établir des relations de parenté. Le principe est de comparer les séquences deux à deux à l'aide de la fonction de comparaison d'*Anagène* afin de construire la matrice des ressemblances entre les régions hypervariables 1 des divers Primates. À partir de là, on peut construire un arbre phylogénétique des Primates similaire à celui-ci obtenu avec *Phylogène*.



NB : Pan corespond au Chimpanzé.

Cet arbre est conforme à celui construit en utilisant d'autres molécules : l'ADN mitochondrial peut donc être utilisé pour reconstruire des phylogénies. De plus, on constate que par rapport aux séquences d'ADN nucléaire utilisées, le nombre de différences entre les séquences d'ADN mitochondrial de l'Homme et des autres Hominidés (Chimpanzé, Bonobo, Gorille) est relativement important, ce qui indique une évolution rapide de cette région de l'ADN mitochondrial. Elle peut donc être utilisée pour résoudre des problèmes de parenté au sein des Homininés.

Le second fichier (Homme moderne, Cro-Magnon, neandertalien, Chimpanzé) peut être utilisé pour traiter la question de la place de l'Homme de Neandertal par rapport à *Homo sapiens* et donc discuter la possibilité ou non d'un métissage entre ces deux *Homo*. Là aussi, il s'agit de comparer les séquences deux à deux afin de construire une matrice des ressemblances et finalement un arbre phylogénétique tel que celui représenté ci-après.



On y voit que les neandertaliens partagent avec *Homo sapiens* un ancêtre commun qui n'est pas celui du Chimpanzé, que les neandertaliens forment un groupe monophylétique à part de celui des Hommes actuels, que les Hommes de Cro-Magnon ne sont pas situés au sein des neandertaliens mais parmi les *Homo sapiens* actuels. Cet ensemble d'informations forme un faisceau d'arguments en faveur de l'absence d'hybridation entre ces deux types d'*Homo* et donc de leur appartenance à deux espèces.

Le troisième fichier des « séquences partielles » présente l'intérêt de fournir des données sur toutes les séquences de neandertaliens ayant été déterminées. Il permet de rechercher si, dans cette courte portion de la région hypervariable 1, les séquences de neandertaliens présentent à certains sites des caractères communs différents de ceux que l'on trouve chez les *Homo sapiens* passés et actuels.

Si on ne veut pas utiliser le deuxième fichier, on peut très bien fournir l'arbre phylogénétique et demander de rechercher en quoi les données du troisième fichier sont en accord avec cet arbre. Il suffit de faire une comparaison avec discontinuité de toutes les séquences partielles.

STABILITÉ ET VARIATION DES GÉNOMES ET ÉVOLUTION

Innovations génétiques – Mutations ponctuelles et filiations entre allèles – *Allèles du gène de l'alpha-antitrypsine*

Informations scientifiques

Cf. page 83 (classe de première)

Pistes d'exploitation pédagogique des données fournies

L'exploitation des données fournies (séquences et documents) dans le thème d'étude sur l'alphaAT permet de bâtir les notions relatives à l'apparition de nouveaux allèles par mutations, le gène de l'alpha AT étant très polymorphe. Les mutations à l'origine de ces allèles sont des substitutions et des délétions d'un nucléotide. Une filiation entre les allèles peut être établie, permettant ainsi de reconstituer une partie de l'histoire évolutive de ce gène. Des documents complémentaires permettent également de discuter de l'origine d'une mutation et de sa diffusion dans les zones géographiques voisines.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Mutations ponctuelles et filiations entre allèles/Allèles du gène de l'alpha-antitrypsine** permet d'atteindre **Allèles AT** qui charge le fichier alleles-AT.edi affichant les séquences nucléiques strictement codantes de quelques allèles du gène de l'alpha AT – allèles M'1, M1, M2, M3, Z, S, NULL 1 et NULL 2

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Mutations ponctuelles et filiations entre allèles/Allèles du gène de l'alpha-antitrypsine** permet de charger les fichiers :

- **InfoAlphaAT.bmp** permettant de définir le phénotype à différents niveaux (clinique, moléculaire et biochimique). Il précise également l'influence de la fumée de cigarette. Il peut donc être utilisé en introduction ou en complément de l'analyse de la filiation des gènes ;
- **AllellesAT.bmp** affichant les informations fournies dans ce tableau qui permettent de relier le phénotype biochimique (concentration plasmatique en alpha AT) et le phénotype clinique pour les différents allèles de l'alpha AT. Les données sur la fréquence des différents allèles permettent également de préciser la notion de polymorphisme ;
- **GenoPhenoAlphaAT.bmp** affichant un document pouvant servir de base à une discussion sur la notion de dominance/récessivité des allèles ;
- **filiationAT.bmp** présentant une activité pédagogique envisageable consistant à proposer cette filiation aux élèves, en leur demandant de la justifier à partir de l'exploitation des séquences fournies (ils doivent identifier et placer sur chaque branche la mutation permettant de la justifier) ;
- **repartitionSZ.bmp** affichant les cartes de répartition qui permettent de discuter de l'origine des allèles S et Z. On constate que la fréquence de l'allèle S est très nettement plus élevée en Espagne, et plus particulièrement en Galice ; de plus, cette fréquence décroît quand on s'éloigne de ces régions. On peut donc supposer que cet allèle est apparu dans cette région. La fréquence la plus élevée de l'allèle « Z » se trouve dans les populations du nord-ouest de l'Europe. On peut supposer que la mutation à l'origine de cet allèle est intervenue dans la lignée germinale d'un individu d'une de ces populations. Ensuite, non seulement l'allèle « Z » a diffusé dans la population nordique, mais a aussi réussi à se répandre dans les populations du sud-est européen.

L'allèle M'1 est considéré comme allèle de référence. La comparaison des autres allèles à cet allèle de référence permet d'aborder les notions relatives aux mutations et au polymorphisme génique. La traduction de ces séquences et la comparaison des séquences protéiques obtenues peuvent permettre de discuter des effets des mutations sur le phénotype moléculaire, biochimique et clinique.

La comparaison des autres allèles avec l'allèle M'1 permet de définir un certain nombre de mutations, substitutions faux-sens, substitutions non-sens, délétions. Cette série d'allèles permet d'introduire la notion de filiation entre allèles : un allèle « b » apparaît par mutation d'un allèle existant « a » dans la lignée germinale d'un individu. Il se répand par la suite plus ou moins dans la population. En mutant chez un autre individu, il

est à l'origine d'un nouvel allèle « c ». Bien entendu, les élèves doivent saisir que les allèles « a » et « b » persistent dans la population.

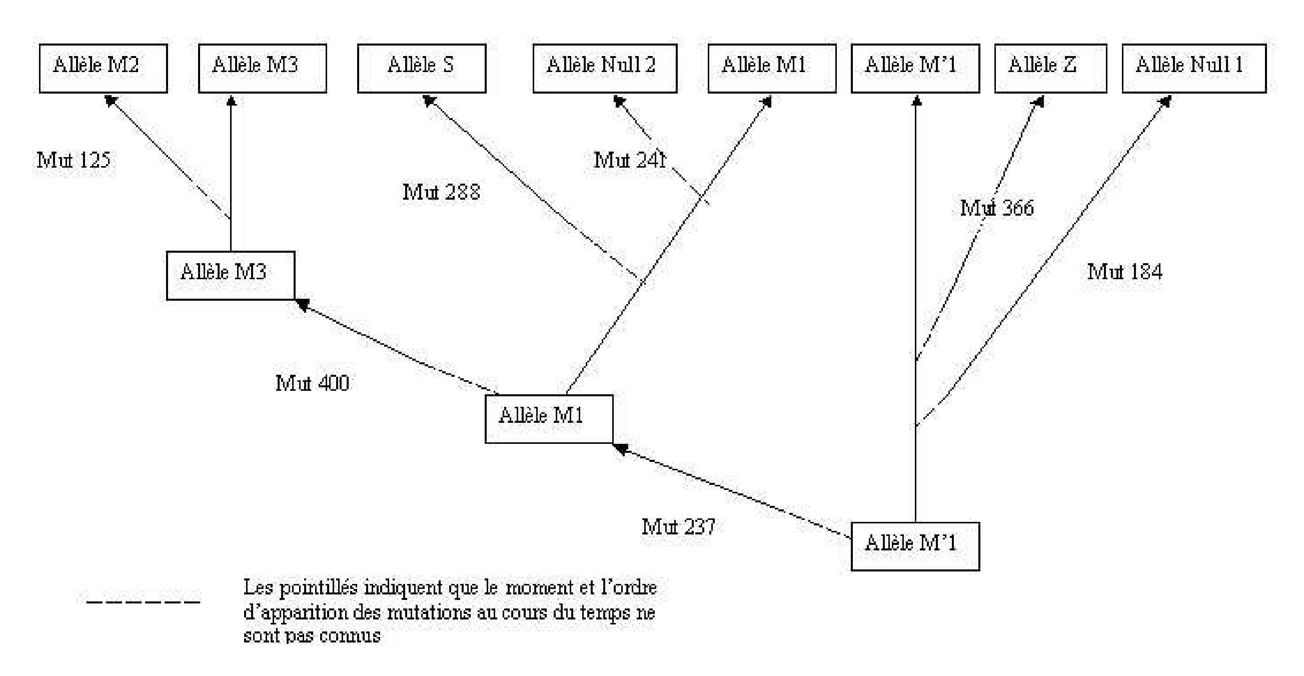
Les allèles « b » et « c » ont en commun la première mutation qui les différencie de l'allèle « a ». L'allèle « c » possède en plus la mutation qui lui est propre. Cela fournit la base de raisonnement que traduit le document sur la filiation entre les allèles de l'alpha-antitrypsine.

La filiation proposée suppose qu'il n'y avait qu'un seul allèle de ce gène dans les premières populations humaines. Si cela est semble-t-il exact pour le gène de l'alpha-antitrypsine, ce n'est pas vrai pour tous les gènes. Ainsi, il est certain que les premières populations humaines étaient polymorphes pour les gènes du système HLA.

Résultats obtenus par l'exploitation des séquences nucléiques fournies

Allèle	**Différence par rapport à l'allèle de référence (M'1)**		**Différence entre la protéine codée par M'1 et la protéine codée par cet allèle**
	Nucléotide	**Codon**	
M1	710 : C → T	Codon 237 : GCG → GTG	A → V
M2	374 : G → A 710 : C → T 1200 : A → C	Codon 125 : CGT → CAT Codon 237 : GCG → GTG Codon 400 : GAA → GAC	R → H A → V E → D
M3	710 : C → T 1200 : A → C	Codon 237 : GCG → GTG Codon 400 : GAA → GAC	A → V E → D
S	710 : C → T 863 : A → T	Codon 237 : GCG → GTG Codon 288 : GAA → GTA	A → V E → V
Z	1096 : G → A	Codon 366 : GAG → AAG	E → K
Null 1	552 : délétion de C	Codon 184 : TAC → TAG	183 acides aminés au lieu de 418
Null 2	710 : C → T 721 : A → T	Codon 237 : GCG → GTG Codon 241 : AAG → TAG	A → V 240 acides aminés au lieu de 418

Tableau de comparaison des allèles de l'alpha AT et des protéines correspondantes



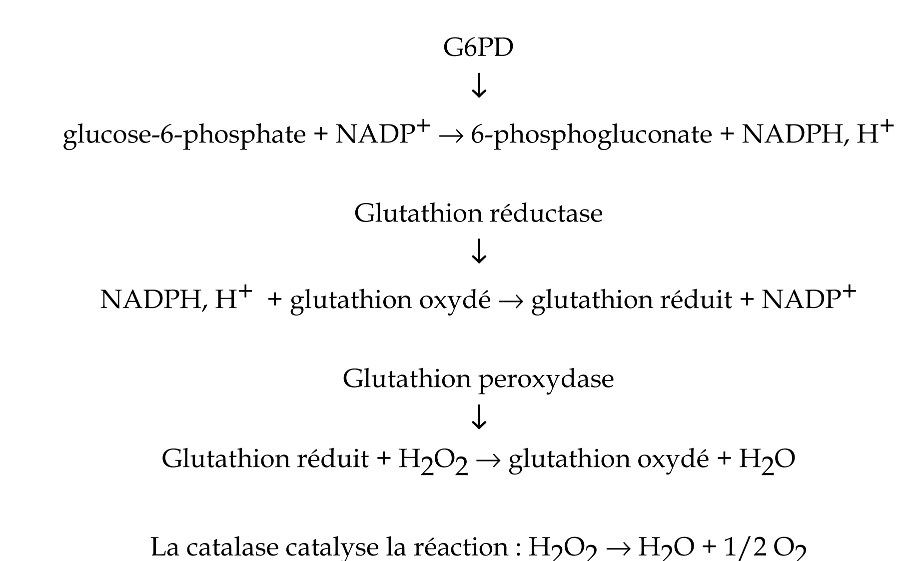
Filiation des allèles de l'alpha AT
Les numéros des mutations se rapportent aux codons et non aux nucléotides

Innovations génétiques – Mutations ponctuelles et filiations entre allèles – *Allèles du gène de la G6PD*

Informations scientifiques

Rôle de l'enzyme G6PD (glucose-6-phosphate déshydrogénase)

C'est une enzyme cytoplasmique présente dans toutes les cellules. Elle catalyse la première réaction de la voie des pentoses phosphates. Cette autre voie du catabolisme glucidique produit du ribose 5 phosphate (qui servira ultérieurement à la synthèse des nucléotides) et du NADPH, coenzyme qui est le principal donneur d'hydrogène dans de nombreuses réactions de biosynthèse. NADPH est aussi indispensable pour que se réalise la destruction du peroxyde d'hydrogène hautement toxique pour la cellule. La chaîne de réaction est la suivante :



Ces quelques réactions nous montrent que pour la destruction du peroxyde d'hydrogène (H_2O_2), la catalase et le glutathion sont indispensables. Avec une G6PD inactive ou très peu active, il y a un arrêt de production de NADPH par la voie des pentoses phosphates. Cela empêche la réduction du glutathion et, par là, la destruction de H_2O_2. On sait d'autre part que le NADPH « stabilise » la catalase. Donc, sans NADPH, H_2O_2 ne sera pas détruit et la cellule sera tuée. Dans les globules rouges, cette situation est d'autant plus dramatique que d'autres enzymes permettant la production de NADPH manquent.

Le polymorphisme du gène G6PD

Le gène codant pour la G6PD est situé dans la partie télomérique du bras long du chromosome X ; il est formé par treize exons et mesure 18 kilo paires de bases environ. Toutefois, sa région codante ne comprend que 1 545 paires de bases, ce qui correspond à une protéine enzymatique formée par 515 acides aminés. On connaît de très nombreux allèles (plus d'une centaine), dont certains ont une fréquence supérieure à 1 %.

Allèles	**Fréquence**			**Activité enzymatique (% par rapport au normal)**	**Manifestations cliniques**
	Afrique	Europe	Méditerranée		
G6pdb	65 %	99,7 %	90-99 %	100	Aucune
G6pda	20 %		< 1 %	85	Aucune
G6pda-1	-	-	-	12	Jaunisse néo-natale ; anémie hémolytique aiguë (médicaments, infection)
G6pda-2	15 %	-	< 12 %	12	Jaunisse néo-natale ; anémie hémolytique aiguë (médicaments, infection)
G6pda-3	-	-	-	12	Jaunisse néo-natale ; anémie hémolytique aiguë (médicaments, infection)
G6pdm	< 0,1 %	-	1 - 8 %	3	Jaunisse néo-natale ; anémie hémolytique aiguë (médicaments, ingestion de fèves, infection)
G6pdseat	-	-	-	25	Rares

Fréquence de quelques allèles de la G6PD contenus dans la banque

La déficience en G6PD

La déficience en G6PD est l'enzymopathie la plus répandue : elle affecterait 400 millions de personnes dans le monde. Les régions les plus touchées sont l'Afrique tropicale, le Moyen-Orient, l'Asie tropicale et subtropicale. Un certain nombre d'allèles codent pour une enzyme G6PD déficiente. La déficience n'est jamais totale : l'absence d'enzyme G6PD est sans doute incompatible avec la vie. Les manifestations cliniques sont la jaunisse néonatale, une anémie hémolytique et, dans des cas sévères, des séquelles neurologiques. Des crises aiguës d'anémie hémolytique peuvent être déclenchées par des infections, des ingestions de fèves et divers médicaments (comme la primaquine). Heureusement, seule une faible proportion des malades déficients en G6PD présentent une anémie hémolytique chronique et, pour les autres, en dehors des crises hémolytiques, il n'y a aucun symptôme particulier. L'action favorisante de l'ingestion de fèves sur le déclenchement des crises hémolytiques est surtout nette chez les personnes possédant l'allèle G6pdm.

Le phénotype des femmes hétérozygotes possédant un allèle « normal » et un allèle « déficient »

Il est classique de considérer comme récessif le phénotype G6PD déficient. En réalité, la situation est plus complexe et le phénotype des femmes hétérozygotes variable, certaines pouvant manifester des signes cliniques de déficience. Cela est en relation avec l'inactivation au hasard d'un des chromosomes X dans chacune des cellules de l'organisme, inactivation qui peut atteindre l'un ou l'autre des chromosomes X. La femme hétérozygote possède deux populations d'hématies, l'une G6PD déficiente, l'autre avec une enzyme G6PD efficace. L'importance relative de ces deux populations varie d'une femme à l'autre.

Pistes d'exploitation pédagogique du gène de la G6PD

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Mutations ponctuelles et filiations entre allèles/Allèles du gène de la G6PD** permet d'atteindre **Allèles G6PD** qui charge le fichier alleles-G6PD-HS.edi affichant les séquences nucléiques strictement codantes de quelques allèles du gène de la G6PD.

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Mutations ponctuelles et filiations entre allèles/Allèles du gène de la G6PD** permet de charger les fichiers :

- **roleG6PD.bmp** affichant les informations suffisantes sur les conséquences de la déficience de l'enzyme ;
- **frequenceG6PD.bmp** indiquant que les allèles G6PDA et G6PDA$^-$ ont une fréquence élevée en Afrique ; à vrai dire, on les trouve aussi dans les Amériques et dans les régions où il y a des populations d'origine africaine. L'allèle G6PDM est ainsi nommé car il est polymorphique dans les pays du bassin méditerranéen, y compris l'Afrique du Nord ;
- **filiationG6PD.bmp** présentant une activité pédagogique envisageable qui propose cette filiation aux élèves, en leur demandant de la justifier à partir de l'exploitation des séquences fournies (ils doivent identifier et placer sur chaque branche la mutation permettant de la justifier).

La comparaison des divers allèles avec l'allèle G6PDB pris comme référence ainsi que l'étude des conséquences des différences sur le polypeptide permet d'établir le tableau suivant.

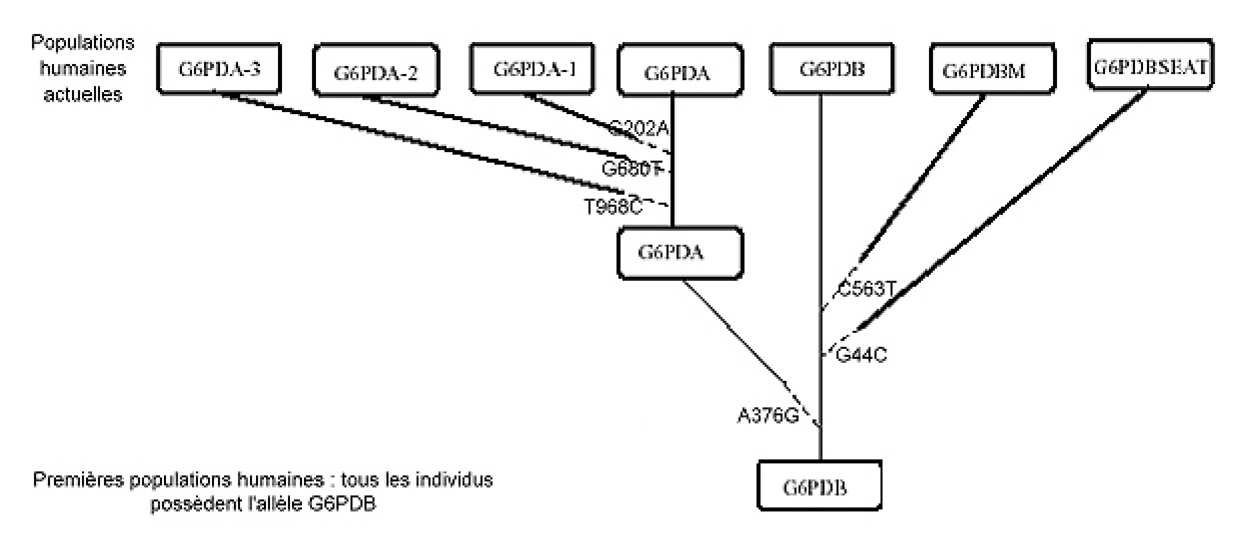
Séquence nucléique			Polypeptide	
Noms des allèles	**Nucléotides changés Nature – Position**	**Codons changés Nature – Position**	**A.A. changés Nature – Position**	**Type de mutation**
G6PDB	(Référence)	(Référence)	(Référence)	
G6PDA	A376 G	AAT126 GAT	Asn126Asp	Substitution faux-sens
G6 PDA -1	A376 G	AAT126 GAT	Asn126Asp	Substitution faux-sens
	G202 A	GTG68 ATG	Val68Met	Substitution faux-sens
G6 PDA -2	A376 G	AAT GAT	Asn126Asp	Substitution faux-sens
	G680 T	CGC227 CTC	Arg227Leu	Substitution faux-sens
G6 PDA -3	A376 G	AAT126 GAT	Asn126Asp	Substitution faux-sens
	T968 C	CTG323 CCG	Leu323Pro	Substitution faux-sens
G6PDM	C563 T	TCC188 TTC	Ser188Phe	Substitution faux sens
G6PDSEAT	G844 C	GAT282 CAT	Asp282His	Substitution faux-sens

Caractéristiques de quelques allèles du gène de la G6PD

Les allèles de ce gène peuvent donc servir de support pour étudier la notion de polymorphisme génique. Certes, les allèles diffèrent uniquement par des substitutions faux-sens et ne permettent pas d'illustrer les diverses modalités des différences alléliques (les allèles résultant de mutations non-sens, d'insertions et délétions décalantes codant pour des protéines tronquées totalement non fonctionnelles sont sans doute rapidement éliminés par la sélection naturelle).

Ce système allélique est aussi un outil pour faire saisir la filiation entre allèles d'un gène.

L'allèle G6PDB, le plus fréquent dans toutes les populations, est sans doute l'allèle ancestral (c'est aussi le plus proche de celui séquencé chez le Chimpanzé). L'allèle G6PDA, répandu en Afrique, résulte d'une substitution au nucléotide 376 de la région codante (A376G) dans l'allèle G6PDB. Les trois allèles G6PDA- présentent cette même différence avec G6PDB plus une autre : tous les trois résultent de mutations survenues au cours de la gamétogenèse d'individus G6PDA et sont donc apparus postérieurement à cet allèle. Par contre, l'allèle G6PDM diffère de G6PDB par une substitution autre que celle trouvée dans G6PDA : il provient d'une mutation intervenue chez un individu G6PDB b mais on ne peut situer chronologiquement sa formation par rapport à G6PDA.



Filiation possible entre les allèles de G6PD. Les pointillés indiquent seulement que le moment et l'ordre d'apparition des allèles par mutation au cours de l'histoire de l'humanité sont inconnus

Innovations génétiques – Mutations ponctuelles et filiations entre allèles – Allèles de la chaîne bêta de l'hémoglobine

Informations scientifiques

Les informations sur la structure de la globine, l'allèle HbC et le phénotype drépanocytaire ont été fournis précédemment. On trouvera ici quelques compléments sur le phénotype thalassémique et les allèles (tous récessifs) qui en sont à l'origine.

Cf. page 51 (classe de première)

Les phénotypes thalassémiques

Il existe en réalité de nombreux phénotypes thalassémiques. Ceux envisagés dans cette banque de données sont en relation avec **l'absence de synthèse d'une chaîne bêta fonctionnelle** de l'hémoglobine. Il s'agit de thalassémies bêta majeures dont les manifestations cliniques débutent dès l'enfance. Les signes hématologiques sont marqués par une anémie hémolytique grave, une morphologie des hématies très irrégulière et par un retard staturo-pondéral et des modifications du squelette en rapport avec l'hémolyse chronique. L'évolution spontanée est constamment mortelle en quelques années. Les malades arrivent à survivre quelques années car l'absence d'hémoglobine A1 fonctionnelle (hémoglobine formée de deux chaînes alpha et de deux chaînes bêta) est partiellement compensée par une synthèse accrue d'hémoglobine A2 (deux chaînes alpha et deux chaînes delta) et la persistance de synthèse d'hémoglobine fœtale (deux chaînes alpha et deux chaînes gamma).

Il n'existe pas de traitement curatif des thalassémies. Les traitements proposés sont basés sur la transfusion mensuelle d'hématies qui, outre qu'elles offrent l'intérêt d'augmenter le taux d'hémoglobine, diminuent l'hyperplasie de la moelle osseuse compensatrice et, par là, préviennent un peu les anomalies du squelette. Ces transfusions fréquentes ont cependant pour complication une surcharge inéluctable en fer dommageable.

Pistes d'exploitation pédagogique des données fournies

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Mutations ponctuelles et filiations entre allèles/Allèles du gène de la globine bêta** permet d'atteindre :

- **Séquences nucléiques** qui charge le fichier GlobineBetaADN.edi affichant les séquences des allèles du gène bêta globine correspondant à trois allèles codant pour une chaîne bêta fonctionnelle (l'allèle HbA étant le plus fréquent), à l'allèle HbS à l'origine du phénotype drépanocytaire et à celles de huit allèles entraînant le phénotype thalassamique ;
- **Séquences protéiques** qui charge le fichier GlobineBetaPRO.edi affichant les séquences protéiques correspondantes.

Le fichier contenant l'ensemble de ces séquences est sans doute le plus approprié pour envisager les divers types de mutations ponctuelles et leurs conséquences sur le phénotype moléculaire. Le principe de la démarche est de :

- comparer les séquences nucléiques des différents allèles du gène de la bêta globine par rapport à l'allèle HbA (allèle considéré comme référence) de façon à identifier les différents types de mutations ponctuelles ;
- comparer les séquences polypeptidiques codées par les divers allèles à la séquence de référence codée par la séquence HbA ;
- mettre les différences au niveau polypeptidique en relation avec celles du niveau nucléique pour faire le point sur les conséquences des mutations sur le phénotype moléculaire.

On aboutit ainsi aux conclusions suivantes :

- certaines mutations (substitution) sont muettes (variant 1) car elles n'ont aucune conséquence sur la séquence du polypeptide du fait de la redondance du code génétique ;
- d'autres substitutions sont neutres car entraînant un changement de la séquence du polypeptide sans en modifier les propriétés (variant 2) ;
- d'autres substitutions sont non-sens car entraînant la synthèse d'un polypeptide non fonctionnel à cause de l'apparition anticipé d'un codon stop ;
- d'autres mutations sont des délétions ou des insertions d'un ou de plusieurs nucléotides entraînant un décalage du cadre de lecture, l'apparition d'un codon stop anticipé et donc la synthèse d'un polypeptide raccourci et non fonctionnel.

Innovations génétiques – Mutations ponctuelles et filiations entre allèles – *Allèles du gène IT15 (chorée de Huntington)*

Informations scientifiques

La chorée de Huntington

Cette maladie se traduit par des mouvements involontaires et désordonnés, des perturbations du psychisme, des pertes de mémoire, des troubles du langage, et finalement de la démence. Ces troubles sont liés à une destruction progressive des neurones cérébraux, notamment de ceux des corps striés, impliqués dans la motricité.

Le gène de la huntingtine (IT 15)

Le gène IT15 est situé sur le bras court du chromosome 14. Il a été isolé et séquencé en 1995 et code pour une protéine, la huntingtine, dont le rôle est actuellement inconnu. Il s'exprime dans de nombreux tissus, notamment les neurones cérébraux des corps striés. Ce gène comprend 67 exons, et sa région codante est caractérisée par la présence d'un triplet CAG (brin non transcrit), près de l'extrémité 5′ (qui correspond à l'extrémité NH2 de la huntingtine), répété plusieurs fois à partir de la position 52.

Dans toutes les populations humaines, il existe de nombreux allèles qui diffèrent par le nombre de répétitions de ce triplet CAG.

On peut regrouper les allèles en deux catégories, en fonction du nombre de répétitions du triplet CAG et du phénotype clinique associé.

Les « variants normaux »

Ces allèles ont un nombre de répétitions du triplet CAG compris entre 9 et 39 ; les allèles les plus fréquents sont ceux qui ont un nombre de répétitions du triplet CAG compris entre 17 et 21. La fréquence de la majorité de ces allèles est supérieure à 1 % : le gène IT15 est donc très polymorphe.

Le tableau ci-dessous indique les fréquences de ces allèles dans une population allemande (d'après *Hum. Mol. Gen.*, 1993, vol. 2, n° 12, p. 2 063-2 067).

Nombre de répétitions du triplet CAG de l'allèle	Fréquence de l'allèle (en %)
11	0,98
12	0
13	0,98
14	0,98
15	1,96
16	2,94
17	20,01
18	13,73
19	8,33
20	10,78
21	14,71
22	6,86
23	5,39
24	2,94
25	3,92
26	0
27	1,47
28	0
29	0,49
30	0,98
31	0,98
33	0,49

Les « variants morbides »

Ils sont à l'origine de la chorée de Huntington, et possèdent un triplet CAG répété entre 36 et 121 fois, avec une fréquence maximale comprise entre 42 et 46 répétitions.

La très grande majorité des personnes atteintes de chorée de Huntington ont un allèle où le triplet CAG est répété plus de quarante fois. Il reste une indétermination pour un triplet répété entre trente-six et trente-neuf fois car selon les cas la possession d'un tel allèle est associée aux signes de la maladie et dans d'autres cas non.

La présence d'une expansion anormalement importante de l'acide aminé glutamine dans la huntingtine conduit à la mort neuronale par l'activation de la machinerie apoptotique : c'est une mort par apoptose (activation d'un programme intrinsèque de mort cellulaire).

Le tableau ci-dessous présente la fréquence des allèles morbides dans une population de personnes atteintes de chorée de Huntington. Toutes les personnes atteintes étaient hétérozygotes (d'après *Hum. Mol. Gen.*, 1993, vol. 2, n° 12, p. 2 063-2 067)

Nombre de répétitions du triplet CAG de l'allèle	Fréquence de l'allèle (en %)
40	6,69
41	11,15
42	11,46
43	10,83
44	12,42
45	12,10
46	8,91
47	7
48	5,10
49	2,23
50	1,91
51	1,59
52	2,87
53	0,96
54	0,64
55	1,27
56	0
57	0,64
58	0,96
61	0,32
63	0,32
73	0,32
75	0,32

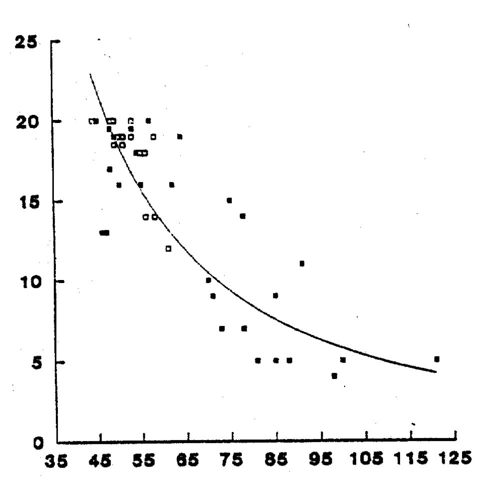
On a constaté l'apparition de cas de chorée de Huntington dans des familles où la maladie n'avait pas été identifiée précédemment. Étant donné la dominance du phénotype morbide, ces cas semblent dus à des néomutations. L'analyse au niveau moléculaire du gène IT15 chez de tels malades et leurs parents a révélé qu'un des parents, très généralement le père, possédait un allèle avec une répétition du triplet CAG comprise entre 30 et 38. Cet allèle paternel avait subi une mutation par expansion au cours de la spermatogenèse conduisant à un allèle où le triplet CAG était répété plus de quarante fois. En outre, dans les familles où le père est atteint, on a constaté une apparition plus précoce de la maladie chez les enfants atteints, liée à la possession d'un allèle morbide ayant plus de triplets CAG que l'allèle paternel. Il semble donc qu'au-delà de trente répétitions, le gène IT15 ait tendance à muter avec expansion du triplet au cours de la spermatogenèse. Cette instabilité n'a pas été retrouvée pour les allèles dont le nombre de répétitions est inférieur à trente. Ainsi, les études au niveau moléculaire de ce gène semblent, pour un certain type d'allèles, révéler une fréquence de mutations assez élevée.

Les relations génotype/phénotype

Chez un individu hétérozygote les deux allèles s'expriment et on trouve donc les deux types de huntingtine dans les cellules. Le phénotype morbide et, donc, par extension, l'allèle morbide, est dominant. L'explication de cette dominance reste très hypothétique, d'autant plus que le rôle physiologique de la huntingtine n'est pas encore élucidé. On pense que la huntingtine avec expansion de la glutamine agit dans le noyau pour induire l'apoptose.

L'âge d'apparition des premiers signes de la maladie varie en fonction de la nature de l'allèle morbide. En règle générale, il est d'autant plus précoce que le nombre de répétitions du triplet CAG est élevé. En revanche, il y a peu de corrélations avec la gravité des symptômes présentés par le patient une fois la maladie exprimée.

En x, nombre de répétitions du triplet CAG dans l'allèle de la personne malade. En y, âge d'apparition des premiers signes de la maladie.



Pistes d'exploitation pédagogique des données fournies

Le gène IT15, dont certains allèles sont responsables de la chorée de Huntington, est particulièrement polyallélique et polymorphe. Bien que le rôle de la protéine codée par ce gène, la huntingtine, soit encore inconnu, les données fournies permettent d'aborder les notions de polymorphisme génique, de mutations par **expansion** d'un motif (triplet) et de **néomutations,** et les relations génotype/phénotype (relations de dominance/récessivité).

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Mutations ponctuelles et filiations entre allèles/Allèles du gène IT15 (Chorée de Huntington)** permet d'atteindre :

- **Allèles IT15** qui charge le fichier alleles-IT15.edi affichant les séquences strictement codantes de quinze allèles du gène IT15, neuf « variants normaux » et six « variants morbides ». Le gène IT15 étant long (9 435 nucléotides en moyenne), on s'est limité, pour chaque allèle, à une portion de la séquence codante, commençant au triplet d'initiation ATG et encadrant la partie répétée (le triplet répété commence à la position 52) ;
- **Famille (Chorée de Huntington)** qui charge le fichier allelesFamillechoree.edi affichant les séquences strictement codantes des allèles du gène IT 15 possédés par les membres de la famille dont l'arbre généalogique est fourni. Le gène IT15 étant long (9 435 nucléotides en moyenne), on s'est limité, pour chaque allèle, à une portion de la séquence codante, commençant au triplet d'initiation ATG et encadrant la partie répétée (le triplet répété commence à la position 52).

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Mutations ponctuelles et filiations entre allèles/Allèles du gène IT15 (chorée de Huntington)** permet d'atteindre :

- **Informations sur la chorée de Huntington** qui charge le fichier choree.bmp présentant des symptômes de la chorée de Huntington et du gène IT15. En introduction, ce document permet de motiver l'étude du polymorphisme de ce gène ;
- **Fréquence de quelques allèles du gène IT15** qui charge le fichier frequenceallelesIT15.bmp affichant les tableaux de fréquence de quelques allèles du gène IT15 dans une population de personnes non malades et dans une population de personnes atteintes de la chorée de Huntington ;
- **Arbre avec sujets atteints de chorée de Huntington** qui charge le fichier arbrechoree.jpg présentant l'arbre généalogique d'une famille touchée par la chorée de Huntington. L'âge d'apparition des premiers symptômes de la maladie des membres atteints de la chorée est indiqué ;
- **Répétition du triplet CAG en fonction de l'âge** qui charge le fichier courbeIT15.jpg

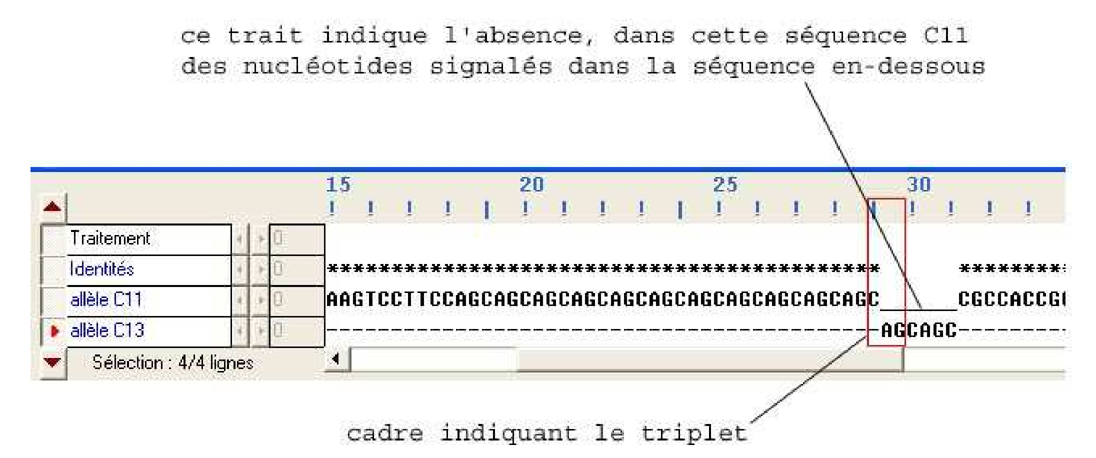
Caractéristiques des allèles du gène IT15

On peut considérer un seul allèle, le C18, par exemple, le traduire et constater qu'il est remarquable par la répétition d'un triplet CAG à partir du codon 18 qui se traduit au niveau protéique par la répétition de l'acide aminé glutamine. Cela facilitera l'analyse des résultats des comparaisons entre allèles. Il faut toutefois noter que cette répétition de la glutamine est aussi suivie d'une répétition de onze prolines codées par différents triplets, alors que la répétition de la glutamine est assurée par le même triplet CAG.

Recherche des différences entre les allèles

Allèles IT15

Après avoir repéré les différences de longueur entre les allèles, il est conseillé de comparer d'abord deux allèles (par exemple les deux premiers) de façon à repérer la répétition d'un triplet CAG à partir de la position 52 et de comprendre que la différence entre les deux allèles consiste en une variation dans le nombre de triplets CAG répétés (un alignement avec discontinuités s'impose, les deux séquences n'ayant pas la même longueur).



Cette illustration montre que l'algorithme d'alignement ne révèle pas que l'expansion de C13 par rapport à C11 commence par le triplet CAG. Cela est dû à la présence d'un C dans le premier triplet qui suit la répétition CAG. Il est utile d'utiliser la règle en triplet pour bien visualiser que la différence débute par un triplet CAG.

NB : Le nom de l'allèle indique le nombre de répétitions du triplet CAG qu'il possède.

Une comparaison de plusieurs allèles peut ensuite être réalisée pour confirmer le fait que les allèles diffèrent uniquement par le nombre de répétitions du triplet CAG. On aboutit donc à l'idée que les différences entre les allèles de ce gène résident dans le nombre de répétitions d'un triplet CAG.

Le document Fréquence de quelques allèles du gène IT15

Il permet alors d'introduire la notion de polymorphisme de ce gène lié à cette répétition de triplets. Il permet également de constater que la chorée de Huntington est liée à des allèles où le nombre de répétitions du triplet CAG est supérieur à quarante.

Conseils pratiques

Il est vraiment conseillé d'utiliser le curseur pour délimiter le triplet, sinon il est difficile à repérer.

Pour compter le nombre de triplets, positionner le grand curseur sur le premier triplet CAG (début au nucléotide 52, ou codon 18), et ensuite cliquer sur la flèche à droite de la barre de défilement ; le défilement se fait alors codon par codon et il suffit de compter le nombre de clics effectués pour connaître le nombre de répétitions du triplet CAG.

Pour dénombrer le nombre de répétitions de triplets, on peut aussi comparer la longueur des différents allèles et diviser la différence par trois.

On peut souhaiter changer le nom des allèles pour éviter que leur nom ne donne la réponse aux élèves sans qu'ils aient à compter eux-mêmes le nombre de répétitions. Procéder alors de la façon suivante : à partir de la fenêtre **Affichage de séquences**, cliquer dans la barre de menus sur **Options**, puis sur **Protéger les données**, ce qui déprotège les données (attention de ne pas indiquer cette procédure aux élèves, sinon, ils peuvent modifier les données d'origine !) ; changer alors le nom de chaque allèle, puis sélectionner les données et les

enregistrer dans la banque de séquences ou dans les thèmes d'étude personnels. Penser à reprotéger les données avant de poursuivre ! (Il suffit de retourner dans **Options** et de cliquer sur **Protéger les données**.)

Ne pas lancer d'alignement avec discontinuité pour un grand nombre de séquences en même temps si la capacité de votre ordinateur est limitée !

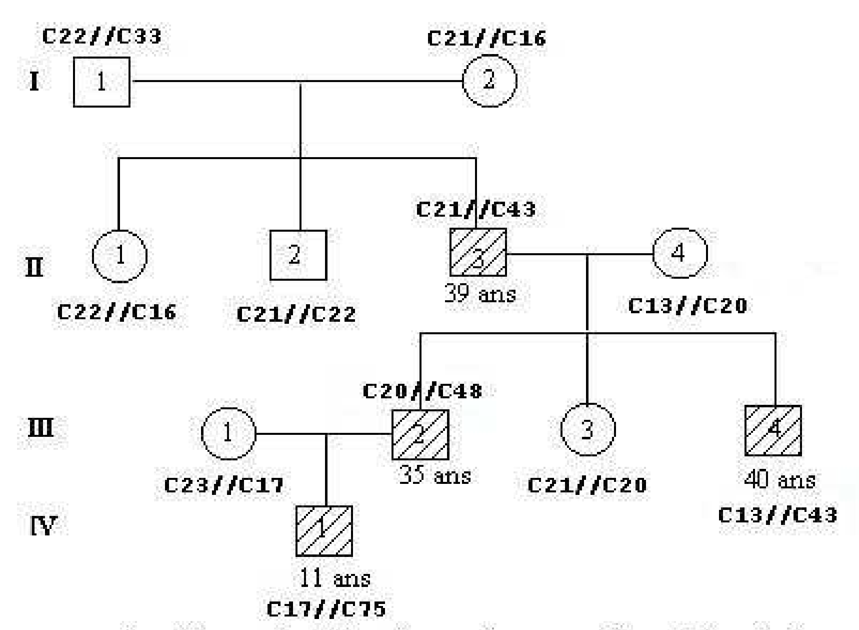
Mise en évidence de néomutations (origine de nouveaux allèles)

La lecture de l'arbre généalogique de la famille confronté avec le document **Répétition du triplet CAG en fonction de l'âge** (fichier courbeITt15.jpg) conduit à suspecter l'apparition de nouveaux allèles au sein de cette famille. La détermination des génotypes permet de tester l'idée.

La comparaison (alignement avec discontinuité) entre les allèles (fichier allelesFamillechoree.edi) d'un individu des générations I, II et III, et l'allèle C21 permet de déterminer le génotype des individus. Pour l'individu IV, 1, il faut utiliser comme références soit l'allèle 21 soit l'allèle 50.

NB : Le choix de ces allèles de référence n'a aucune signification biologique. Compte tenu de l'algorithme d'alignement, les allèles C21 et C50 sont ceux qui visualisent le mieux l'expansion de chacun des allèles des individus de la famille.

Allèles des individus	Nombre de répétitions du triplet CAG	Allèle du gène IT15
I1 allèle 1	22	C22
I1 allèle 2	33	C33
I2 allèle 1	21	C21
I2 allèle 2	16	C16
II1 allèle 1	22	C22
II1 allèle 2	16	C16
II2 allèle 1	21	C21
II2 allèle 2	22	C22
II3 allèle 1	21	C21
II3 allèle 2	43	C43
II4 allèle 1	13	C13
II4 allèle 2	20	C20
III1 allèle 1	23	C23
III1 allèle 2	17	C17
III2 allèle 1	20	C20
III2 allèle 2	48	C48
III3 allèle 1	21	C21
III3 allèle 2	20	C20
III4 allèle 1	13	C13
III4 allèle 2	43	C43
IV1 allèle 1	17	C17
IV1 allèle 2	75	C75



Arbre généalogique d'une famille avec individus atteints de chorée

Les génotypes des individus atteints de chorée indiquent que les allèles à l'origine de cette maladie sont dominants. En conséquence, il y a une contradiction entre les phénotypes des individus de la première génération et le génotype de l'individu II3. Celui-ci possède l'allèle C21 de sa mère, mais aucun allèle de son père. Cela conduit à l'idée que l'allèle C43 de II3 provient d'une mutation subie par un des allèles du père lors de la spermatogenèse (sans doute l'allèle C33). Cette idée de néomutation permet également l'interprétation du génotype de l'individu IV1.

Innovations génétiques – Duplications et familles multigéniques – *Gènes des globines*

Informations scientifiques

Cf. page 124 (classe terminale)

L'exploitation des données fournies permet de bâtir les notions relatives à l'apparition de nouveaux gènes par duplication génique suivie de mutations. Les gènes de la famille des globines codent pour des protéines ayant conservé globalement la même fonction (fixation de dioxygène grâce au groupement hème). La disposition des différents gènes de cette famille sur deux chromosomes différents permet d'aborder le phénomène de transposition (ou translocation) génique. Quelques données paléontologiques sont utilisables pour dater approximativement certaines duplications géniques.

Pistes d'utilisation pédagogique

Les gènes des globines forment une famille multigénique

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Duplications et familles multigéniques/Gènes des globines** permet d'atteindre :

- **Nucléiques** qui charge le fichier genes-glob-humaines.edi affichant les séquences nucléiques représentant la partie strictement codante (du codon d'initiation au codon stop) des gènes de globines humaines (alpha 1, alpha 2, bêta, gamma A, gamma G, zêta, delta, epsilon, thêta, myoglobine) ;
- **Protéiques** qui charge le fichier pro-glob-humaines.edi affichant les séquences protéiques des globines humaines (alpha 1, alpha 2, bêta, gamma A, gamma G, zêta, delta, epsilon, thêta, myoglobine).

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Duplications et familles multigéniques/Gènes des globines** permet d'atteindre :

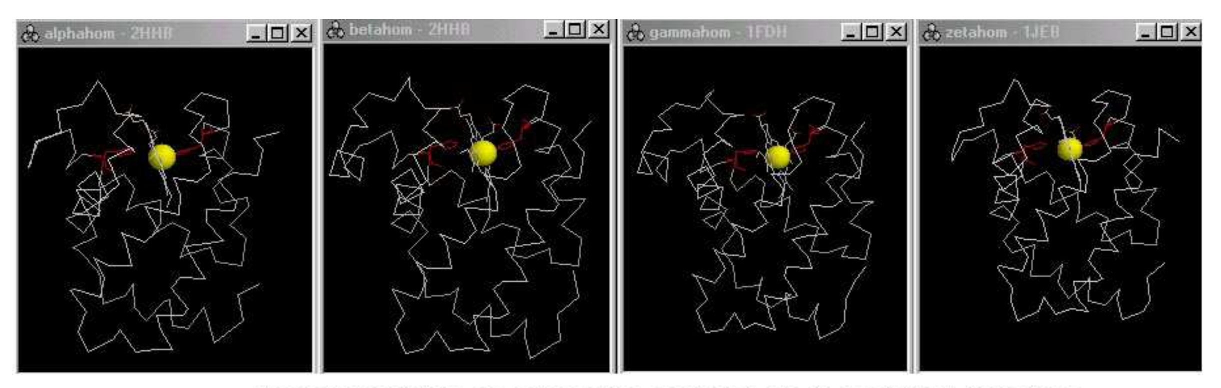
- **Présentation des globines humaines** qui charge le fichier globineshumaines.bmp affichant un petit texte précisant l'existence de plusieurs globines humaines, accompagné d'un graphique exprimant, en pourcentage, la synthèse de ces différentes globines en fonction de l'âge. Peut servir à poser le problème de l'existence de plusieurs globines humaines ayant le même rôle, et qui sont produites à différents moments de la vie de l'être humain ;
- **Localisation chromosomique** qui charge le fichier chrombetaglob.jpg affichant un schéma de localisation des gènes de globines humaines, y compris les pseudogènes sur les chromosomes 11 (gènes du groupe bêta) et 16 (gènes du groupe alpha). Ce document peut être important pour bien faire comprendre qu'il s'agit de gènes différents et non d'allèles d'un même gène. L'existence de deux familles de gènes localisées sur deux chromosomes différents peut aussi suggérer une parenté plus grande entre les gènes du groupe bêta ou entre les gènes du groupe alpha qu'entre les gènes de deux groupes différents. Utilisé après avoir posé la notion de duplication génique, ce document est utile pour introduire la notion de transposition d'un gène ;
- **Comparaison des globines en 3D** qui charge le fichier comp3Dglobines.jpg affichant les représentations en squelette carboné et en rubans de quatre globines humaines : les globines alpha, bêta, gamma et zêta. Ces structures ont été obtenues avec le logiciel *RasTop*. La similitude de structure des globines présentées peut suggérer l'idée d'une parenté entre ces molécules, donc entre les gènes qui les codent ;
- **Filiation globines** qui charge le fichier arbreglobines.jpg révélant la phylogénie qui traduit les relations de parenté entre l'ensemble des différents gènes de globines humaine (myoglobine, famille alpha, famille bêta). Des indications de dates sont portées sur cette figure ; elles situent approximativement dans le temps les principales duplications géniques ; elles ont été obtenues en combinant deux sortes d'information : les différentes globines présentes chez différents taxons et les dates d'apparition de ces taxons ;
- **Datation des duplications** qui charge le fichier datadupli.jpg rappelant sous forme d'un tableau les dates d'apparition des principales classes de Vertébrés, ainsi que des indications permettant de dater les duplications alpha/bêta et bêta/delta.

Fichiers des molécules en 3D

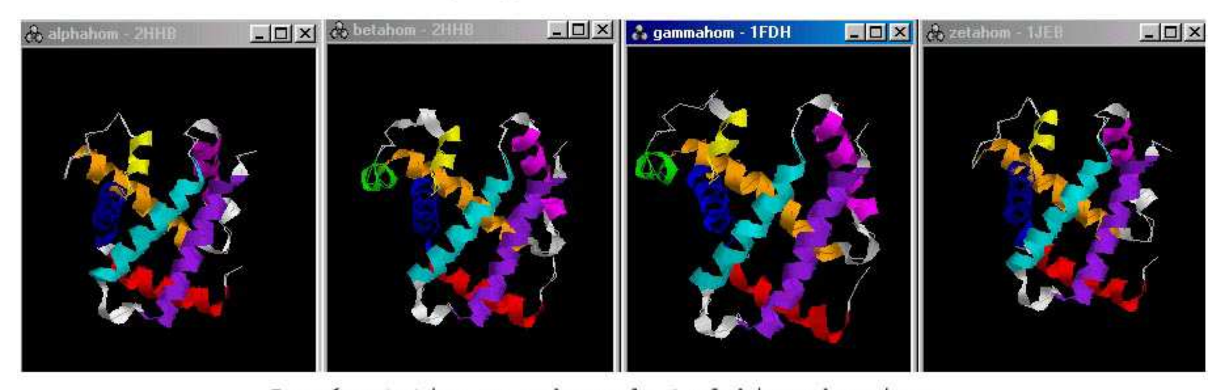
Dans le dossier 3D sous Anagene2, le répertoire **FamilleGlobines** contient les fichiers .pdb suivants : alphahom.pdb, betahom.pdb, gammahom.pdb et zetahom.pdb.

Des similitudes de structure et de fonction pour les globines humaines

Toutes les globines humaines, globines alpha, bêta, zêta, gamma G, gamma A, epsilon présentent des structures qui se ressemblent : de telles similitudes confortent l'idée d'une origine commune.



Représentation en squelette carboné de 4 globines humaines
(le groupement hème est mis en évidence)



Représentation en rubans de 4 globines humaines
(les secteurs en hélice sont mis en évidence de façon colorée)

Des similitudes de séquences protéiques et nucléiques

Les séquences nucléiques ou les séquences peptidiques des globines humaines présentent un degré de similitude qui ne peut être dû au hasard. Les globines sont donc des protéines homologues qui dérivent donc d'un même gène ancestral (de la même façon, on observe des similitudes dans l'organisation de ces différents gènes : ils possèdent tous trois exons et deux introns et ont des longueurs proches). Les gènes des globines forment donc une famille multigénique.

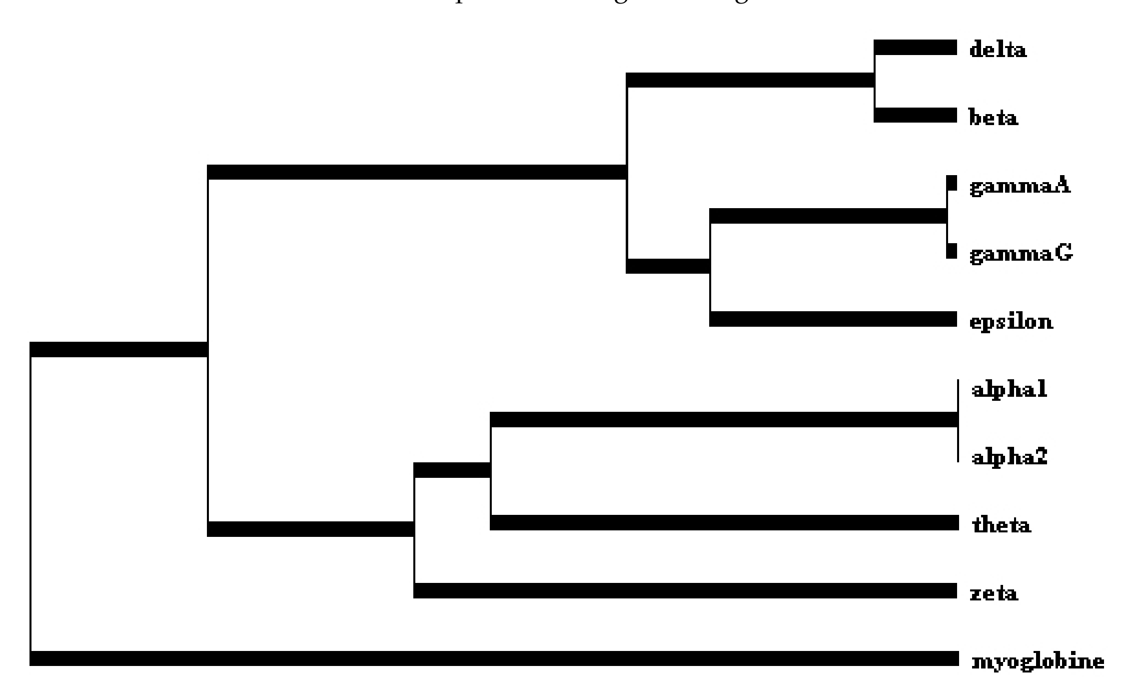
Afin d'établir l'arbre de filiation des gènes de cette famille de globines humaines, on relève les simulitudes entre les séquences nucléiques de ces gènes, ce qui permet de construire une matrice :

	Alpha1	Alpha2	Bêta	GammaA	GammaG	Delta	Epsilon	Thêta	Zêta	Myoglobine
Alpha1	100									
Alpha2	99,5	100								
Bêta	57	56,8	100							
GammaA	54,5	54,1	75,9	100						
GammaG	54,5	54,1	76,1	99,8	100					
Delta	56,3	55,9	92,6	76,1	76,4	100				
Epsilon	54,5	54,3	78,8	83,6	83,8	78,6	100			
Thêta	73,2	73	52,7	53,8	53,8	52,4	53,1	100		
Zêta	68,8	68,5	55	55,9	55,9	54,5	56,9	64,8	100	
Myoglobine	40,9	40,6	40,2	42,2	41,9	40,2	41,9	41,1	40,9	100

Matrice des identités (en %) entre les globines humaines obtenue à partir d'un alignement avec discontinuité des séquences codantes des gènes pour ces protéines

NB : Ces résultats ont été obtenus en choisissant une séquence de référence (la première) et en maintenant toutes les autres dans l'ordre alphabétique des noms. Du fait de l'algorithme de comparaison, ces résultats peuvent être légèrement différents de ceux obtenus par comparaison des séquences deux à deux.

On déduit de cette matrice l'arbre d'évolution probable des gènes des globines humaines :



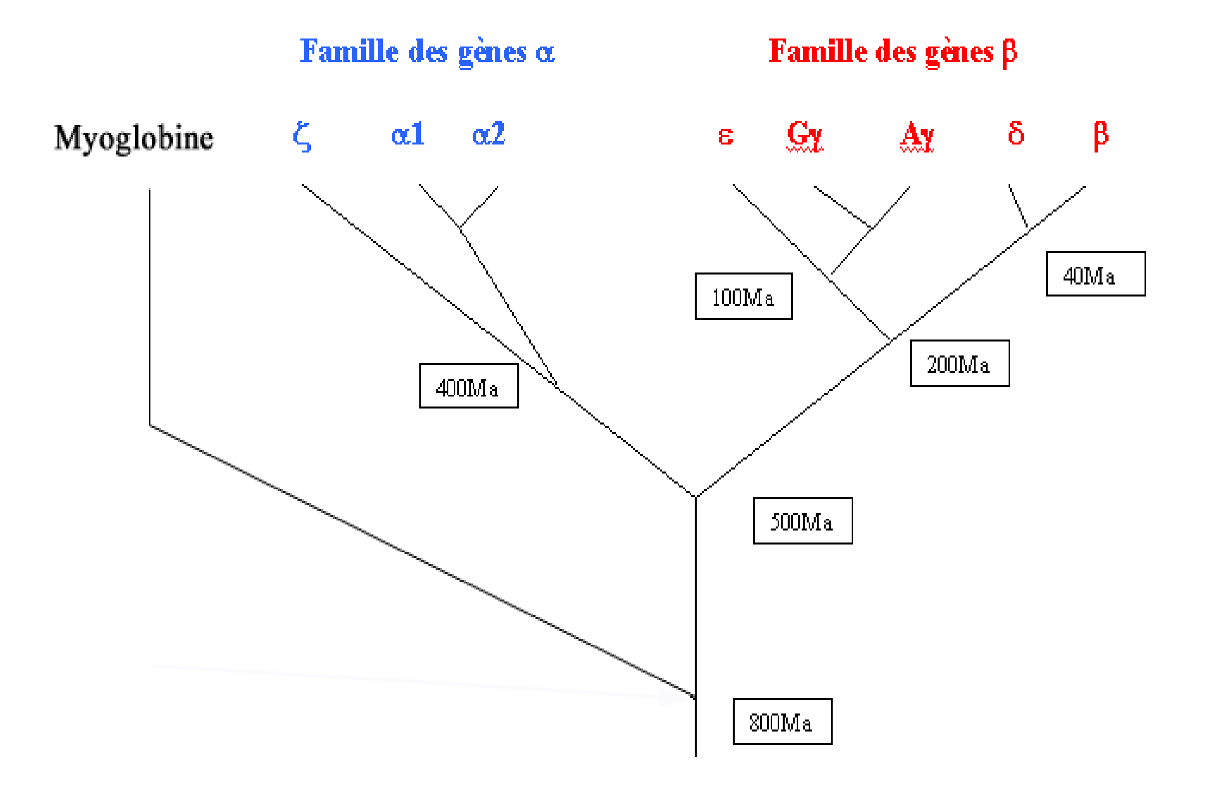
Arbre de filiation des globines humaines obtenu avec le logiciel Phylogène

Les mécanismes à l'origine de cette famille multigénique

Chaque nœud de l'arbre obtenu précédemment correspond à une duplication génique ; sur chaque branche, des mutations différentes ont été fixées (les mutations apparaissant au hasard), responsables d'une évolution indépendante des deux duplicatas.

Un phénomène de translocation est à faire intervenir pour expliquer la localisation des gènes sur des chromosomes différents. Cette translocation (ou transposition) a dû accompagner la duplication à l'origine des deux familles de gènes (famille alpha et famille bêta).

Tous les Vertébrés, à l'exception des Agnathes, ayant deux gènes de globine (alpha et bêta) et les plus anciens poissons étant datés d'environ 450 mA alors que les premiers Vertébrés sont apparus il y a environ 550 mA, on en déduit que la première duplication a dû avoir lieu entre ces deux dates.



La famille des globines est donc issue de duplications suivies de mutations qui amènent à une diversification des gènes.

Il est possible de proposer divers exercices utilisant les documents sur l'arbre d'évolution des gènes de globine et sur les dates d'apparition des principales classes de Vertébrés. Par exemple, l'arbre indique une date de 200 millions d'années pour la duplication gamma–bêta, bien antérieure à l'apparition des différents ordres de Mammifères. On peut demander aux élèves de tester cette donnée à l'aide des séquences des globines bêta et gamma humaines et de la globine bêta du Chien. La comparaison de ces séquences montre que les similitudes entre la globine bêta humaine et la globine bêta du Chien sont plus importantes qu'entre les globines bêta et gamma humaines. Cela signifie que la divergence Primates/Carnivores est postérieure à la duplication bêta-gamma. L'ancêtre commun à l'Homme et au Chien possédait déjà deux gènes bêta et gamma. C'est un caractère ancestral, la duplication ayant eu lieu bien avant, au tout début de l'histoire des Mammifères placentaires.

Innovations génétiques – Duplications et familles multigéniques – *Gènes des opsines*

Informations scientifiques

Cf. page 129 (classe terminale)

Pistes d'exploitation pédagogique

L'exploitation des données fournies permet d'aborder la notion d'apparition de nouveaux gènes par duplication génique suivie de mutations différentes. Les nouveaux gènes codent ici pour des protéines qui, tout en ayant conservé la même fonction globale (absorption de radiations lumineuses) présentent des propriétés légèrement différentes (les longueurs d'ondes absorbées par les différents pigments ne sont pas les mêmes).

Plusieurs arguments suggèrent l'idée d'une origine commune pour les gènes des opsines :

- les opsines présentent de nombreuses similitudes : leurs propriétés et leurs structures se ressemblent. Elles sont toutes impliquées dans la conversion de la lumière en mouvement ionique puis en signal nerveux. Chaque molécule photosensible contient sept hélices alpha transmembranaires ;
- les séquences d'acides aminés des opsines absorbant le vert et des opsines absorbant le rouge sont particulièrement semblables. Trois résidus localisés près du rétinal déterminent la différence spectrale entre les récepteurs pour le vert et pour le rouge. Le remplacement d'un résidu polaire par un résidu non polaire (par exemple, la sérine par l'alanine) à chacune de ces positions déplace λ_{max} vers le rouge d'environ 10 nm. Trois acides aminés expliquent la majeure partie de la différence de 30 nm entre les maxima d'absorption ;
- les séquences nucléotidiques ou les séquences peptidiques des opsines humaines présentent des similitudes importantes.

La famille multigénique des opsines

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Duplications et familles multigéniques/Gènes des opsines** permet d'atteindre :

- **Nucléiques** qui charge le fichier genes-Opsines.edi affichant les séquences nucléiques strictement codantes des gènes des opsines humaines rouge, verte et bleue ;

- **Protéiques** qui charge le fichier pro-opsines.edi affichant les séquences protéiques des opsines rouge, verte et bleue humaines.

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Duplications et familles multigéniques/Gènes des opsines** permet d'atteindre :

- **Informations opsines** qui charge le fichier presopsines.jpg affichant un texte de présentation des opsines humaines (incluant le spectre d'absorption) et précisions sur les opsines présentes chez les autres Primates. Ce document peut être utilisé pour poser le problème (existence, chez un même organisme, de plusieurs molécules aux rôles très voisins). Il est aussi utilisable pour dater la dernière duplication génique à partir des informations concernant les opsines présentes chez les autres Primates ;

- **Spectres d'absorption des opsines** qui charge le fichier spectropsines.jpg présentant le spectre d'absorption des opsines humaines ;

- **Comparaison structure 3D opsines** qui charge le fichier comparopsines.jpg révélant la comparaison des structures 3D (représentation en rubans) des trois ospines humaines (document obtenu avec le logiciel *RasTop*). Les similitudes constatées sont un argument en faveur d'une origine commune de ces protéines, donc des gènes qui les codent ;

- **Localisation chromosomique** qui charge le fichier chromopsines.jpg montrant la localisation chromosomique des gènes des opsines humaines. Une fois vue leur origine commune et les mécanismes à l'origine de leur apparition, ce document permet de discuter de la plasticité importante du génome, en abordant notamment le phénomène de transposition de gènes ;

- **Phylogénie des opsines** qui charge le fichier phyloopsines.jpg montrant la phylogénie des gènes des opsines humaines (sans légende). Ce document peut être complété de façon à y placer les mécanismes intervenant dans l'histoire de ces gènes (duplications et fixations de mutations indépendamment dans les deux branches).

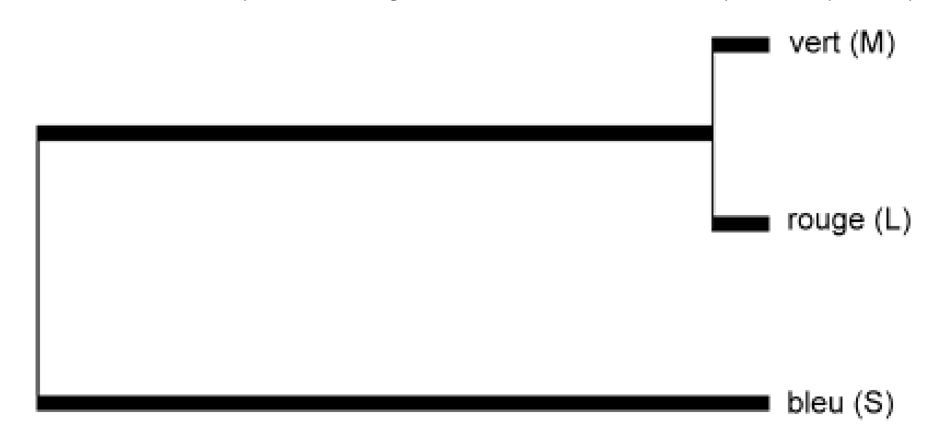
La comparaison de séquences protéiques et/ou nucléiques permet de dégager la notion de molécules homologues (protéines homologues ou gènes homologues) grâce à l'importance des similitudes constatées. Le relevé des différences permet de préciser le degré de parenté entre les protéines et/ou les gènes, et donc d'établir une phylogénie pour ces gènes.

	Gène de l'opsine rouge (L)	Gène de l'opsine verte (M)	Gène de l'opsine bleue (S)
Gène de l'opsine rouge	100 %		
Gène de l'opsine verte	97,7 %	100 %	
Gène de l'opsine bleue	57,1 %	57,6 %	100 %

Matrice des identités créée à partir d'un alignement avec discontinuité des gènes des opsines

	Opsine « rouge » (L)	Opsine « verte » (M)	Opsine « bleue » (S)
Opsine rouge	100 %		
Opsine verte	95,9 %	100 %	
Opsine bleue	42 %	43,1 %	100 %

Matrice des identités créée à partir d'un alignement avec discontinuité des opsines (séquences protéiques)



Arbre phylogénétique obtenu à partir de l'exploitation des matrices des identités

La phylogénie obtenue permet de discuter des mécanismes à l'origine de l'apparition de gènes différents à partir d'un seul.

Datation des duplications

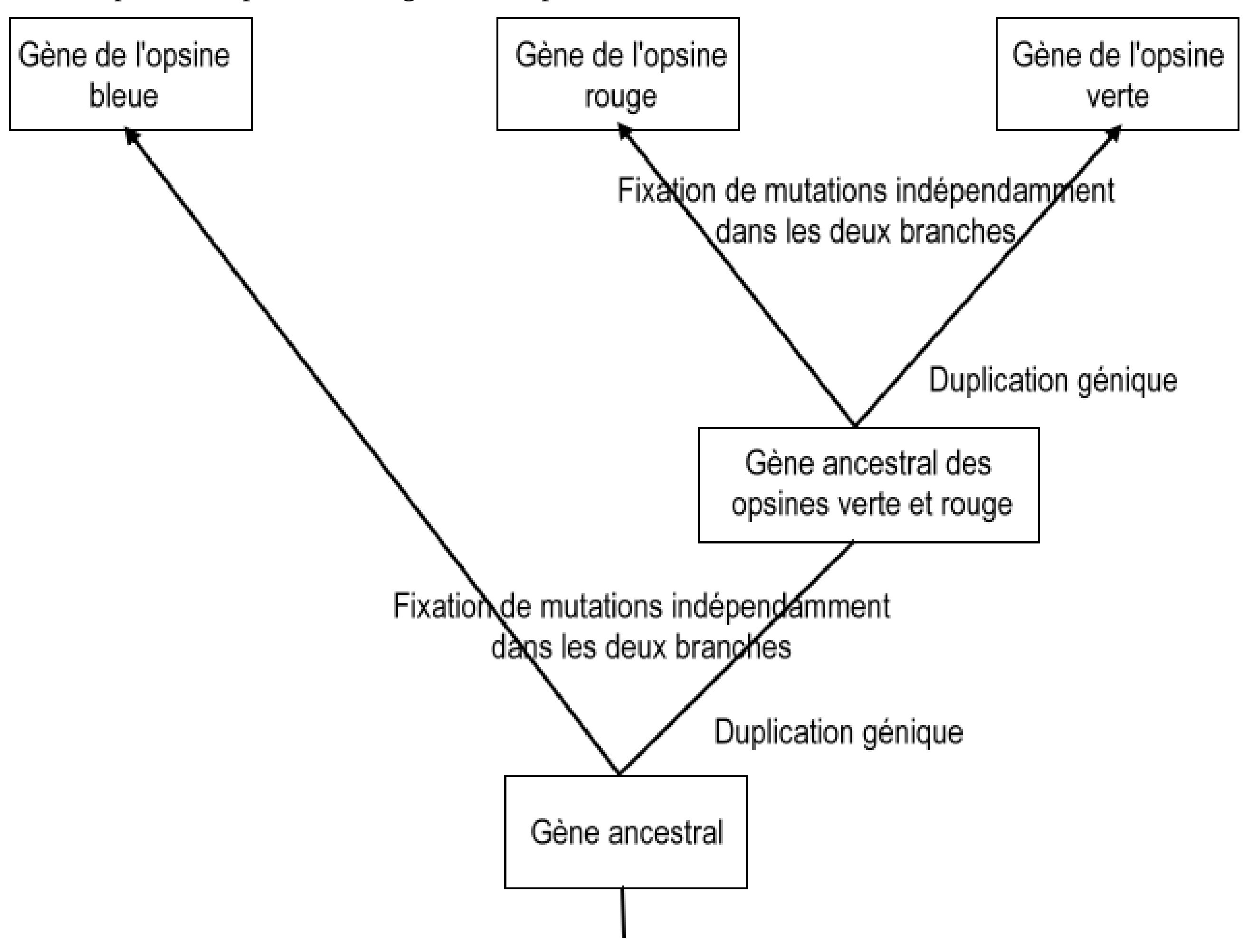
La prise en compte d'observations réalisées chez d'autres Primates permet de dater approximativement la dernière duplication génique : seuls les Singes de l'Ancien Monde possèdent trois gènes d'opsine et notamment un gène d'opsine rouge et un gène d'opsine verte. Les Singes du Nouveau Monde ne possédant que deux gènes, on peut dire que la duplication à l'origine des gènes des opsines verte et rouge à partir d'un même gène ancestral a dû avoir lieu dans la lignée menant aux Singes de l'Ancien Monde, et a donc dû se produire après la séparation de cette lignée de celle des Singes du Nouveau Monde, soit il y a environ 40 à 23 Ma.

Translocation de gènes

Les gènes des opsines provenant tous d'un même gène ancestral, mais étant situés sur deux chromosomes différents, il faut imaginer un phénomène de translocation d'un des deux exemplaires de gènes lors de la duplication du gène ancestral.

L'histoire évolutive de la famille des opsines

On peut donc demander aux élèves de compléter l'arbre phylogénétique à l'aide des informations fournies par la comparaison des gènes des opsines comme ci-dessous :



Reconstitution de l'histoire de la famille des opsines

Innovations génétiques – Duplications et familles multigéniques – *Gènes des hormones hypophysaires et placentaires*

Informations scientifiques

Les hormones antéhypophysaires et placentaires forment deux familles multigéniques : la famille des gènes LH, FSH, TSH et HCG et la famille des gènes GH, HPRL et HLP.

La famille des gènes LH, FSH, TSH et HCG

Les hormones antéhypophysaires LH, FSH et TSH et l'hormone HCG sont des glycoprotéines formées par l'assemblage de deux chaînes polypeptidiques, les chaînes alpha et bêta. Ces quatre hormones diffèrent uniquement par leur chaîne bêta qui est la chaîne responsable de la spécificité fonctionnelle de ces hormones.

Ces hormones jouent des rôles différents dans l'organisme : les hormones LH et FSH participent au contrôle hormonal du fonctionnement de l'appareil reproducteur chez l'homme et chez la femme, l'hormone TSH stimule le fonctionnement thyroïdien et l'hormone HCG stimule le corps jaune lors d'un début de grossesse.

Ces chaînes bêta sont codées par des gènes qui sont situés sur les chromosomes différents (chromosome 1 pour le gène de TSH bêta, chromosome 11 pour le gène de FSH bêta et chromosome 19 pour les gènes de LH bêta et de HCG bêta).

Plusieurs arguments permettent de dire que ces gènes forment une famille multigénique :

- ces quatre gènes présentent la même structure globale (trois exons et deux introns) ;
- ces quatre gènes et les protéines pour lesquelles ils codent présentent des similitudes de séquences assez importantes ;
- au niveau protéique, on peut constater la présence de douze résidus cystéine à des sites identiques pour les quatre hormones. Or, ces résidus, en permettant la formation des ponts disulfures, jouent un rôle essentiel dans la stabilisation de la structure tridimensionnelle de ces chaînes et donc de leurs propriétés.

On peut donc considérer que les similitudes constatées au niveau protéique et nucléique traduisent une origine commune pour toutes ces hormones et les gènes qui les codent. Ces gènes sont issus d'un même gène ancestral, et ils sont apparus par duplications géniques suivies de mutations indépendantes. La duplication la plus récente est celle qui est à l'origine des gènes LH bêta et HCG bêta. Ce qu'il est important de noter ici, c'est que la divergence **des copies dupliquées a conduit à des gènes codant pour des polypeptides ayant des fonctions différentes**.

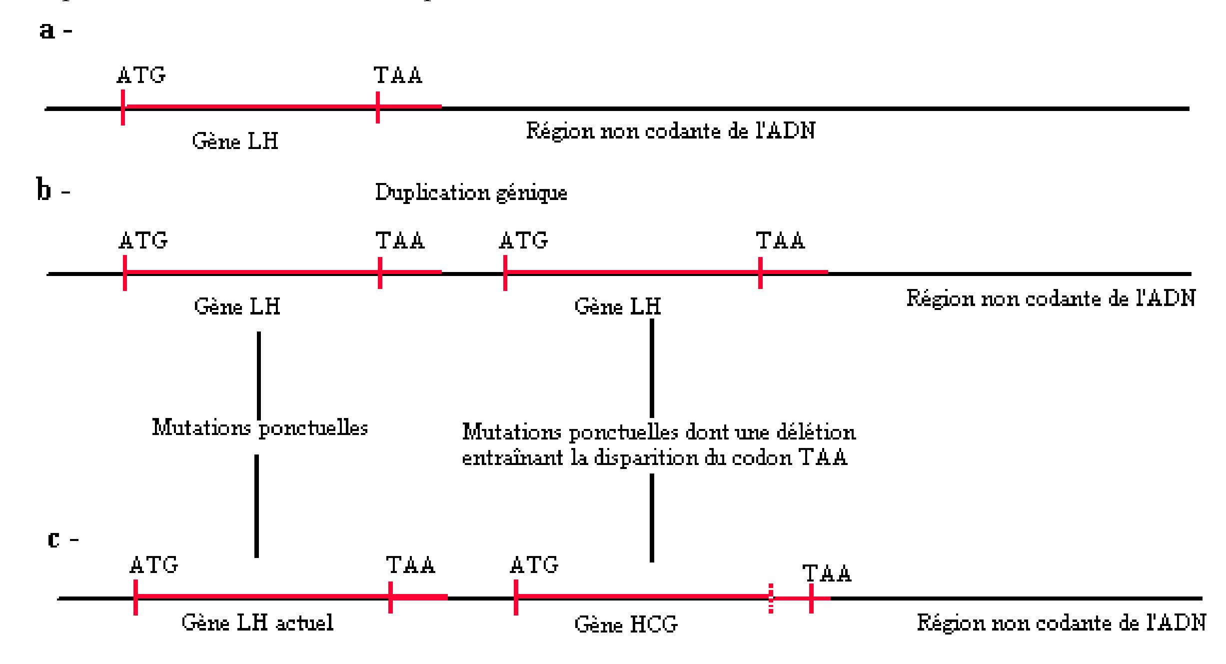
Tous les Vertébrés possèdent le gène de LH bêta, mais seuls les Mammifères placentaires (notamment les Primates et les Équidés) possèdent le gène de HCG bêta. On peut donc penser que le gène HCG bêta provient d'un gène LH bêta ancestral, la duplication étant postérieure à l'apparition des Mammifères placentaires.

La comparaison des protéines LH bêta et HCG bêta montre que ces deux protéines n'ont pas la même longueur : 165 acides aminés pour HCG bêta et 141 acides aminés pour LH bêta (méthionine comprise). Or, la comparaison des séquences nucléiques des gènes de ces deux hormones montre une délétion d'un nucléotide (en position 402) ; cette délétion entraîne un décalage du cadre de lecture lors de la traduction, décalage qui fait disparaître le codon stop présent dans le gène de LH (codon n° 142). C'est cette mutation par délétion qui est donc responsable de l'allongement de la protéine HCG bêta par rapport à la protéine LH bêta.

Remarque

La duplication du gène ne concerne pas uniquement la séquence strictement codante mais aussi celle qui se trouve en amont et en aval de celle-ci. Pour faciliter la compréhension de l'allongement de la séquence codante de HCG (et donc de sa protéine) par rapport à LH, nous avons fourni une séquence appelée LHADNC constituée par la séquence codante de LH, suivie d'une séquence de nucléotides homologue de la fin de celle de HCG mais non transcrite dans LH. La comparaison de LHADNC et de HCG permet d'expliquer que la protéine HCG a 24 acides aminés de plus que LH : par suite de la délétion, le gène HCG incorpore dans sa séquence codante la région en aval de LH non transcrite.

On peut conduire les élèves à la représentation suivante :



a - Situation initiale il y a plus de 60 Ma dans une espèce ancestrale des Primates.
b – Duplication génique de LH bêta.
c - Situation actuelle : la délétion dans le duplicata, en entraînant un décalage du cadre de lecture, a fait disparaître le codon stop. La région codante du gène incorpore une partie de l'ADN non codant situé en aval, jusqu'à un nouveau codon stop.
En toute rigueur, pour bien comprendre tout cela, il faut savoir que l'ARNm comporte une région normalement non codante après le codon stop UAA. Si ce codon stop disparaît, l'ARNm est traduit jusqu'à la rencontre d'un nouveau codon stop.

La famille des gènes GH, HPRL et HLP

Les hormones GH (hormone de croissance produite par l'antéhypophyse), HPRL (hormone prolactine, produite par l'antéhypophyse) et HLP (hormone lactogène placentaire produite par le placenta) ont des actions différentes mais apparentées. La HLP agit sur la glande mammaire en stimulant la synthèse protéique, comme la HPRL (toutefois plus efficace), et la GH stimule également la synthèse protéique dans de nombreux tissus.

Ces hormones sont codées par des gènes situés sur des chromosomes différents : chromosome 6 pour le gène de la HLP, chromosome 17 pour les gènes de la GH et la HPRL.

Les similitudes de séquences nucléiques et protéiques constatées permettent de dire que ces gènes forment une famille multigénique :

Ces trois gènes ont donc une origine commune et sont issus d'un même gène ancestral par duplications géniques suivies de mutations différentes.

La duplication la plus récente est celle qui est à l'origine des gènes GH et HLP.

Tous les Vertébrés tétrapodes possèdent les gènes GH et HPRL, mais seuls les Mammifères placentaires possèdent le gène HLP.

On peut donc considérer qu'une première duplication (à l'origine des gènes GH et HPRL) a eu lieu chez les premiers tétrapodes, avant que n'apparaissent les Amphibiens, et qu'une deuxième duplication a eu lieu beaucoup plus récemment, il y a environ 100 Ma.

Pistes d'exploitation pédagogique des données fournies

L'exploitation des données fournies sur les hormones antéhypophysaires et placentaires permet de dégager les notions liées aux mécanismes d'évolution du génome, et notamment à l'apparition de nouveaux gènes par duplication génique suivie de mutations indépendantes.

Ici, **les mutations qui sont apparues ont amené à des fonctionnalités nouvelles** pour les protéines codées puisque toutes ces hormones ont des rôles différents. Enfin, les gènes codant pour ces hormones étant situés sur des chromosomes différents, on illustre bien avec l'étude de cet exemple la souplesse du génome au cours de l'évolution.

Les gènes des hormones étudiées ici constituent deux familles multigéniques : la famille des gènes LH, FSH, TSH et HCG, et la famille des gènes de la GH, HPRL et HLP.

La comparaison des séquences protéiques et nucléiques au sein de chacun des groupes proposés (famille des gènes LH, FSH, TSH, HCG et famille des gènes GH, HPRL et HLP) permet de mettre en évidence des similitudes moléculaires et de dégager la notion de familles multigéniques. La notion de duplication génique suivie de mutations indépendantes permet alors d'expliquer l'apparition de nouveaux gènes à partir d'un même gène ancestral.

La prise en compte du taux d'identités permet de préciser les relations de parenté au sein de chacune de ces familles. Les données paléontologiques fournies permettent quant à elles de comprendre comment on peut dater approximativement certaines duplications géniques.

Enfin, les données fournies peuvent servir de support à une réflexion sur les effets des mutations et sur leur conservation :

- l'étude des gènes et protéines LH et HCG permet de montrer comment une mutation peut, en supprimant un codon stop, entraîner un allongement de la protéine synthétisée, ce qui a contribué, avec l'apparition d'autres mutations, à l'obtention d'une protéine aux fonctionnalités nouvelles ;
- l'alignement multiple des séquences protéiques des hormones LH, FSH, HCG et TSH permet de repérer l'existence de sites identiques remarquables, en particulier les cystéines. Or, ces résidus, en permettant la formation de ponts disulfures, jouent un rôle essentiel dans la stabilisation de la structure tridimensionnelle de la protéine et donc dans ses propriétés. On peut alors discuter du fait que ces sites soient très conservés d'une hormone à l'autre : une mutation à leur niveau doit arriver (aussi fréquemment qu'ailleurs dans la molécule), mais ne se fixe pas car elle entraîne une non-fonctionnalité de la protéine, donc nuit à la survie de l'individu porteur qui ne peut alors pas la transmettre.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Duplications et familles multigéniques** permet d'atteindre :

- **Gènes GH, HPRL et HLP** pour accéder à :

- **Nucléiques** qui charge le fichier genes-familleGH.edi affichant les séquences nucléiques strictement codantes (du codon d'initiation au codon stop) des gènes de GH (hormone de croissance produite par l'antéhypophyse), HPRL (hormone prolactine produite par l'antéhypophyse), HLP (hormone lactogène placentaire produite par le placenta) ;

- **Protéiques** qui charge le fichier hormones-familleGH.edi affichant les séquences protéiques des hormones GH (hormone de croissance produite par l'antéhypophyse), HPRL (hormone prolactine produite par l'antéhypophyse), HLP (hormone lactogène placentaire produite par le placenta).

- Gènes des hormones FSH, LH, TSH et HCG pour accéder à :

- **Nucléiques** qui charge le fichier genes-familleLH.edi affichant les séquences nucléiques strictement codantes (du codon d'initiation au codon stop) des gènes de LH (hormone lutéinique produite par l'antéhypophyse), FSH (*Folliculate Stimulating Hormon* produite par l'antéhypophyse), TSH (*Thyroïd Stimulating Hormon*, produite par l'antéhypophyse) et HCG (hormone gonadotrophine chorionique, produite par les villosités choriales) ;

- **Protéiques** qui charge le fichier hormones-familleLH.edi affichant les séquences protéiques des hormones LH (hormone lutéinique produite par l'antéhypophyse), FSH (*Folliculate Stimulating Hormon* produite par l'antéhypophyse), TSH (*Thyroïd Stimulating Hormon*, produite par l'antéhypophyse) et HCG (hormone gonadotrophine chorionique, produite par les villosités choriales).

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Duplications et familles multigéniques** permet d'atteindre **Gènes des hormones FSH, LH, TSH et HCG** et **Gènes GH, HPRL et HLP** pour accéder à :

- **Information hormones** qui charge le fichier infoshormones.bmp présentant, pour chacune des hormones considérées ici, la localisation chromosomique du gène, le rôle, les cellules productrices et les groupes de Vertébrés qui la synthétisent. Ce document peut être utilisé pour discuter des phénomènes de transposition de gène lors de la duplication, de l'acquisition de nouvelles fonctionnalités pour les protéines codées par les nouveaux gènes produits et de la datation approximative de certaines duplications géniques ;
- **Duplication LH HCG** qui charge le fichier dupliLH_HCG.jpg et **Duplication GH HLP** qui charge le fichier dupliGH_HLP.jpg : ces documents peuvent servir de base de discussion pour poser la notion de duplication génique ; ils peuvent être à compléter par les élèves ou à justifier à partir des analyses de séquences effectuées ;
- **Phylogénie famille GH** qui charge le fichier phylofamilleGH.jpg résumant l'histoire évolutive de la famille des gènes GH/HPRL/HLP et qui peut être complété par les élèves ou à justifier à partir des analyses de séquences effectuées. Une autre utilisation possible de ces documents consisterait à les considérer comme des modèles de référence de l'histoire d'une famille multigénique donnée, et à demander aux élèves en quoi ils sont confortés par l'analyse des séquences à leur disposition.

Comparaisons obtenues pour la famille des gènes LH, FSH, TSH et HCG

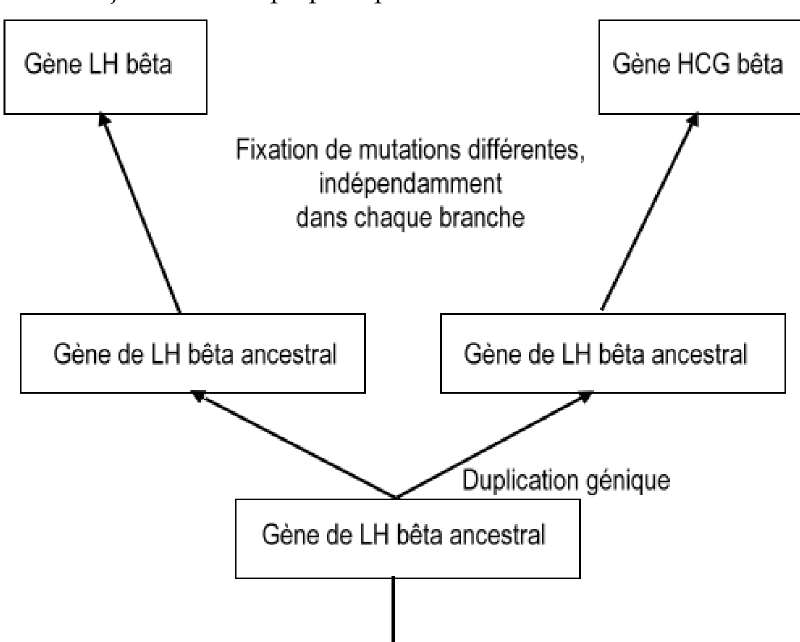
	LH bêta	HCG bêta	FSH bêta	TSH bêta
LH bêta	100 %			
HCG bêta	79,1 %	100 %		
FSH bêta	54,4 %	54,4 %	100 %	
TSH bêta	45,5 %	46,5 %	45,8 %	100 %

Matrice des identités obtenue à partir d'un alignement avec discontinuité des gènes de ces hormones

	LH bêta	HCG bêta	FSH bêta	TSH bêta
LH bêta	100 %			
HCG bêta	71,5 %	100 %		
FSH bêta	34,1 % %	33,3 %	100 %	
TSH bêta	39,1 % %	34,8 %	35,5 %	100 %

Matrice des identités obtenue à partir d'un alignement avec discontinuité des séquences protéiques de ces hormones

Le plus fort taux d'identité est entre les gènes LH bêta et HCG bêta (ou entre les hormones elles-mêmes). On en déduit que ces deux gènes dérivent d'un gène ancestral récent, par un phénomène de duplication génique ; les différences qu'ils présentent aujourd'hui s'expliquent par la fixation de mutations différentes dans les deux cas :



La LH bêta étant présente chez tous les Vertébrés alors que la HCG bêta n'étant présente que chez quelques ordres de Mammifères placentaires (Primates, Équidés), on peut dater approximativement la duplication à l'origine des gènes de ces deux hormones : elle est postérieure à l'apparition des Mammifères placentaires. Les Équidés n'étant pas les Mammifères les plus étroitement apparentés aux Primates, on peut supposer que l'apparition de HCG résulte d'une évolution convergente dans les lignées des Primates et des Équidés. D'autre part, cela permet de dire que le gène ancestral devait coder pour une molécule de type LH bêta.

La mise en relation des mutations dans les gènes LH bêta et HCG bêta avec les différences observées au niveau protéique permet de discuter de l'effet de certaines mutations : ici, une mutation par délétion en position 402 apparue dans le gène HCG bêta provoque la disparition d'un codon stop et donc l'allongement de la protéine synthétisée (la HCG bêta possède 165 acides aminés, la LH bêta n'en possède que 141).

Les taux d'identité entre les gènes FSH bêta, TSH bêta et LH bêta/HCG bêta sont suffisants pour dire que ces gènes sont homologues (et que ces protéines sont homologues), mais les différences ne sont pas assez significatives pour pouvoir réaliser une phylogénie pertinente (entre 54,4 % et 45,5 %, la différence n'est pas assez nette pour pouvoir vraiment distinguer sur une phylogénie les duplications à l'origine de ces gènes, d'un point de vue chronologique). Comme on obtient quatre gènes à partir d'un gène ancestral unique, on peut dire qu'il y a eu trois duplications dans cette famille de gènes.

Les gènes des hormones FSH et TSH n'étant pas sur le même chromosome que les gènes LH et HCG, il faut également faire intervenir un phénomène de transposition (ou translocation) de gènes lors de chacune des deux duplications les plus anciennes. Les quatre hormones n'ayant pas les mêmes rôles, on peut discuter de l'effet des mutations apparues et fixées différemment selon les branches ; ces mutations sont à l'origine des fonctionnalités nouvelles acquises par ces hormones.

Comparaisons obtenues pour la famille des gènes GH, HPRL et HLP

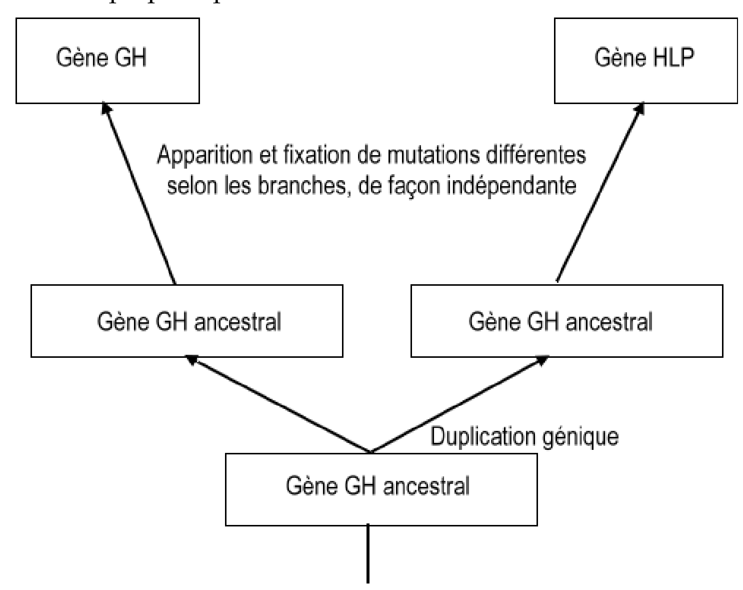
	GH	HPRL	HLP
GH	100 %		
HPRL	44,3 %	100 %	
HLP	92,7 %	47,1 %	100 %

Matrice des identités obtenue à partir d'un alignement avec discontinuité des gènes de ces hormones

	GH	HPRL	HLP
GH	100 %		
HPRL	22,9 %	100 %	
HLP	85,3 %	24,4 %	100 %

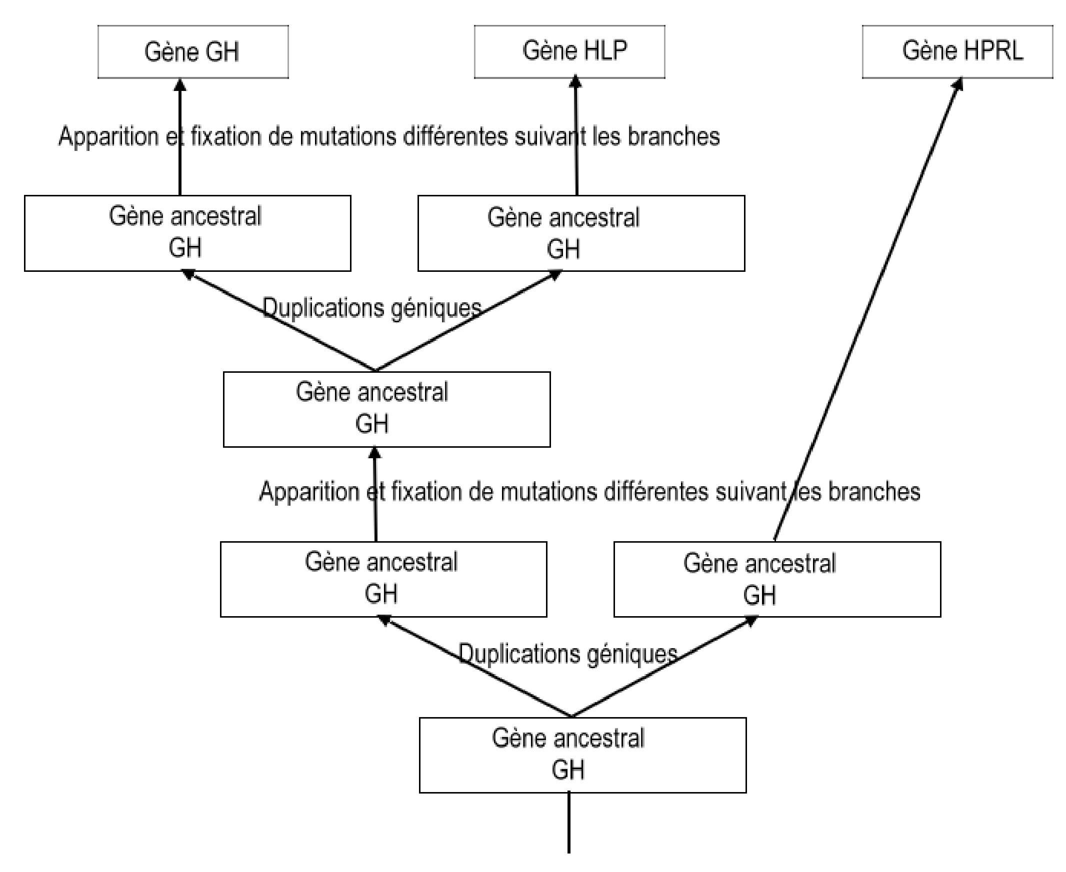
Matrice des identités obtenue à partir d'un alignement avec discontinuité des séquences protéiques de ces hormones

Le plus fort taux d'identité est entre les gènes GH et HLP (ou entre les hormones elles-mêmes). On en déduit que ces deux gènes dérivent d'un gène ancestral récent, par un phénomène de duplication génique ; les différences qu'ils présentent aujourd'hui s'expliquent par la fixation de mutations différentes dans les deux cas :



La GH étant présente chez tous les Vertébrés alors que la HLP n'est présente que chez les Mammifères placentaires, on peut dater approximativement la duplication à l'origine des gènes de ces deux hormones chez l'Homme : elle a eu lieu chez une espèce à l'origine de tous les placentaires actuels, soit environ entre 85 et 100 Ma. D'autre part, cela permet de dire que le gène ancestral devait coder pour une molécule de type GH.

Les taux d'identité entre les gènes HLP et GH/HPRL sont beaucoup plus faibles, mais suffisants pour dire que ces gènes sont homologues (et que ces protéines sont homologues), surtout si l'on considère la conservation de plusieurs acides aminés à des positions précises dans les protéines codées et la structure globale identique de ces gènes. Comme on obtient trois gènes à partir d'un gène ancestral unique, on peut dire qu'il y a eu deux duplications dans cette famille de gènes.



Les gènes des hormones GH et HLP n'étant pas sur le même chromosome que le gène HPRL, il faut également faire intervenir un phénomène de transposition (ou translocation) de gènes lors de la duplication la plus ancienne (tous les Vertébrés possédant les gènes GH et HPRL, on peut dire que la duplication à l'origine de ces deux gènes est plus ancienne que l'apparition du groupe des Amphibiens, soit il y a environ 400 Ma).

Les trois hormones n'ayant pas les mêmes rôles, on peut discuter de l'effet des mutations apparues et fixées différemment selon les branches ; ces mutations sont à l'origine des fonctionnalités nouvelles acquises par ces hormones.

Innovations génétiques – Duplications et familles multigéniques – Gènes homéotiques

Informations scientifiques

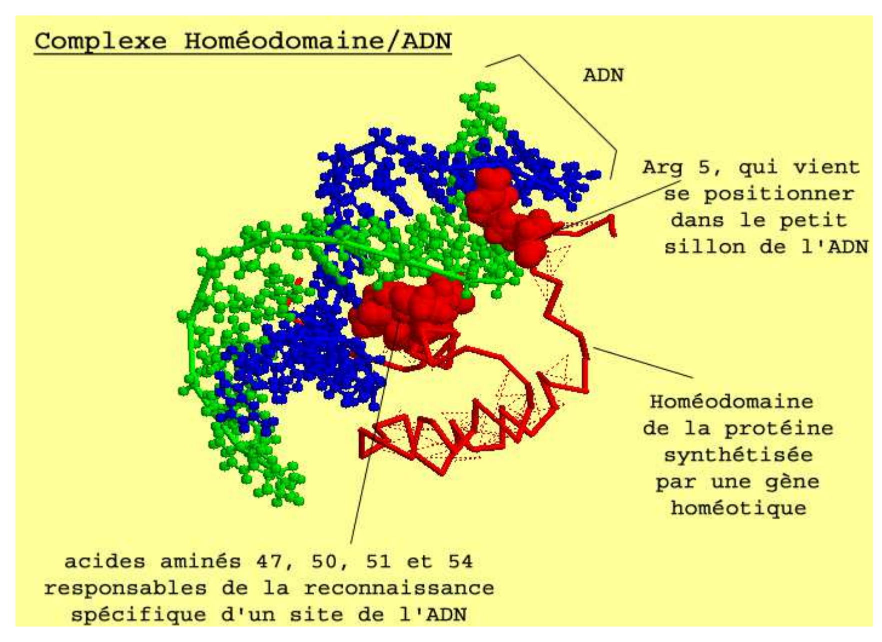
Cf. page 45 (classe de seconde)

Les gènes homéotiques interviennent au cours du développement embryonnaire, notamment pour diriger l'expression d'autres gènes et permettre ainsi la mise en place des différents organes ; on les qualifie de gènes « maîtres » ou gènes « architectes ». Ils déterminent la disposition des organes le long de l'axe antéro-postérieur.

Les gènes homéotiques sont des gènes régulateurs (activation ou inhibition) de l'expression d'autres gènes.

Ce sont des gènes à homéoboîte (ou homeobox), c'est-à-dire qu'ils codent pour une protéine dont un domaine est très conservé au cours de l'évolution. Ce domaine protéique, appelé homéodomaine, a été découvert en 1983 ; il comporte soixante acides aminés et sa structure secondaire comporte trois hélices formant le motif HLH (Helix Loop Helix). Ce domaine se fixe à l'ADN, ce qui contribue à ouvrir localement l'ADN pour permettre la transcription.

La figure ci-dessous a été obtenue avec le logiciel *RasTop.*



Les gènes homéotiques sont organisés en complexes homéotiques chez les animaux.

Les complexes homéotiques

La disposition et l'organisation des gènes homéotiques sont bien connues chez deux organismes : la Drosophile et la Souris.

Chez la Drosophile, les gènes homéotiques forment le complexe Hom-C porté par le chromosome 3 ; ce complexe est composé de deux groupes de gènes : le groupe Antennapedia et le groupe Bithorax, chacun de ces groupes comprenant plusieurs gènes :

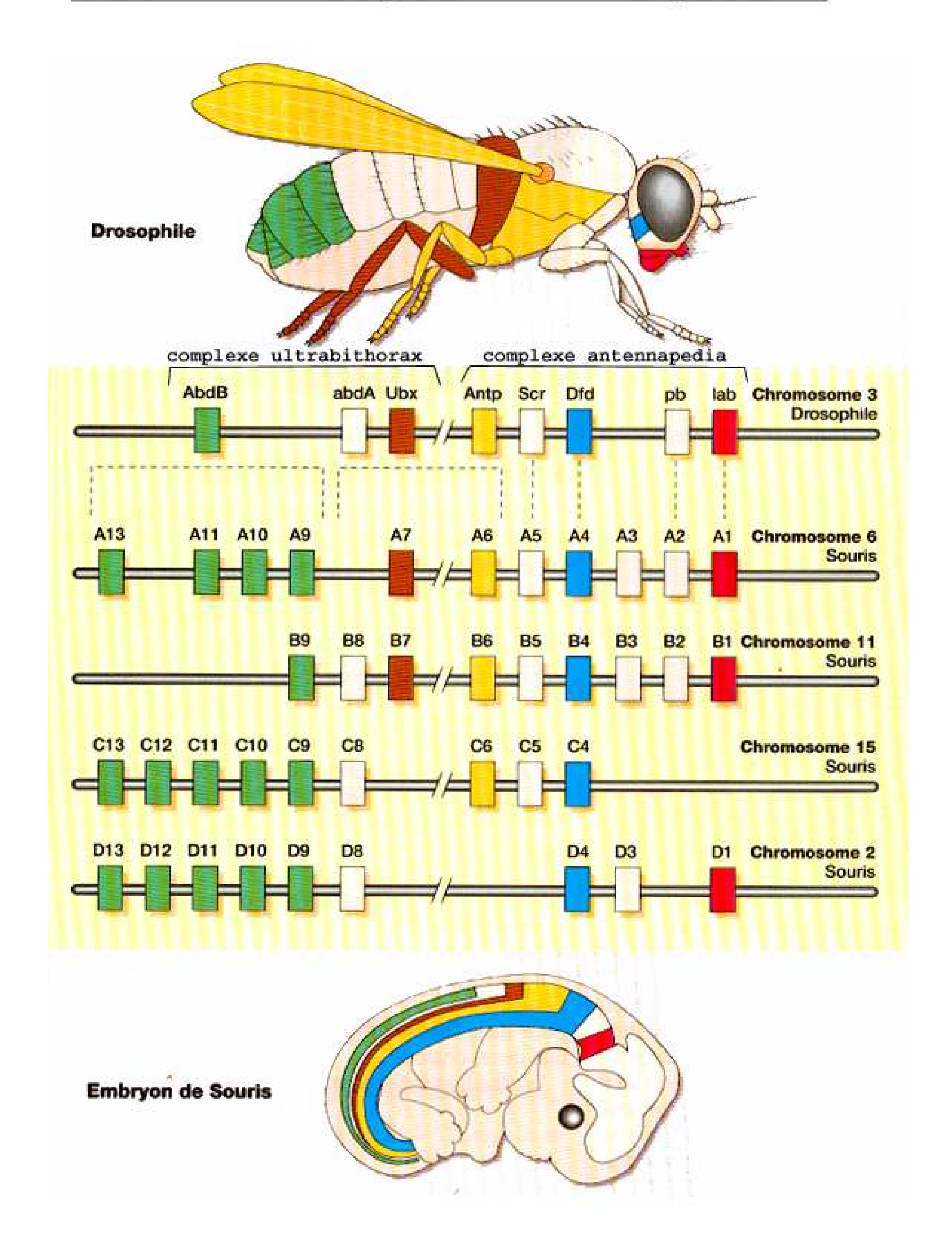
- Antennapedia : gènes lab (labial), Antennapedia (antp), Sex comb reduced (Scr), Deformed (dfd), proboscipedia (pb) ;
- Bithorax : gènes ultrabithorax (ubx), abdominal A (AbdA) et abdominal B (AbdB).

Chez la Souris, les gènes homéotiques forment quatre complexes Hox répartis sur quatre chromosomes :

- Complexe des gènes Hox A sur le chromosome 6 ;
- Complexe des gènes Hox B sur le chromosome 11 ;
- Complexe des gènes Hox C sur le chromosome 15 ;
- Complexe des gènes Hox D sur le chromosome 2.

Il y a colinéarité entre la disposition des gènes homéotiques sur les chromosomes et les régions du corps dans lesquelles ils s'expriment : en parcourant l'ADN de 3' vers 5', on trouve des gènes dont les lieux d'action s'échelonnent de l'avant vers l'arrière de l'animal.

Disposition des gènes homéotiques et secteurs d'expression correspondants



Lors de l'embryogenèse, la combinaison des produits de plusieurs gènes Hox donne une identité relative aux cellules embryonnaires le long de l'axe antéro-postérieur. Chaque cellule embryonnaire est affectée d'une valeur positionnelle, qui résulte d'une combinatoire de plusieurs gènes Hox.

Pistes d'exploitation pédagogique

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Duplications et familles multigéniques/Gènes homéotiques** permet d'atteindre :

- **Nucléiques** qui charge le fichier BoxFamillesHomme.edi affichant les séquences de gènes chez l'Homme et la Drosophile ;
- **Protéiques** qui charge le fichier DomFamillesHomme.edi affichant les séquences protéiques correspondantes.

Les gènes homéotiques des Arthropodes et des Vertébrés forment une famille multigénique

Au sein d'un même organisme, par exemple la Souris, tous les gènes homéotiques présentent une similitude importante de séquence au niveau de l'homéoboîte : ce sont donc des gènes homologues, cette homologie témoignant de leur origine commune. De plus, l'ordre des gènes sur chaque chromosome (donc au sein d'un même complexe) est similaire d'un complexe à l'autre. On peut donc dire que l'on a des gènes paralogues (exemple : HoxA4, HoxB4, HoxC4 et HoxD4 sont paralogues).

Comparaisons entre les gènes homéotiques des Vertébrés et des Arthropodes :

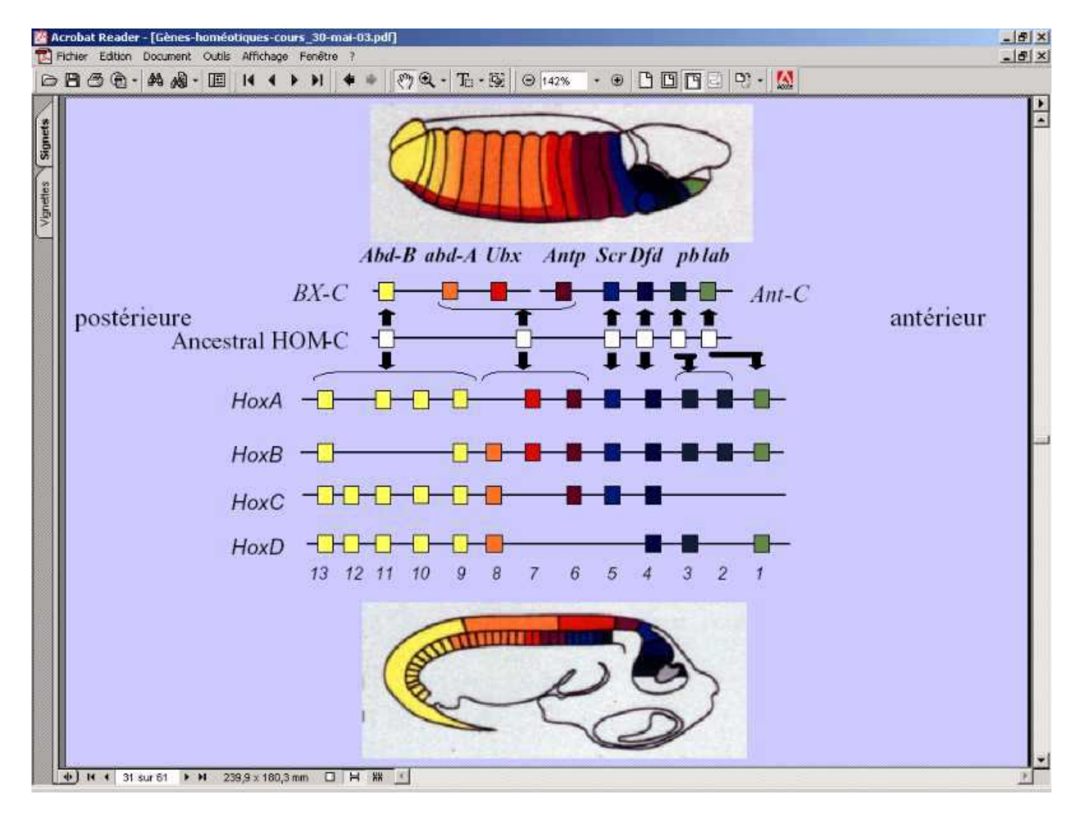
- les gènes des complexes Hox de Souris et Hom-C de Drosophile présentent des similitudes dans leur ordre de disposition sur les chromosomes et, dans les deux cas, ils sont impliqués dans le positionnement et l'identité cellulaire le long de l'axe antéro-postérieur ;
- des expériences de transgenèse interspécifique ont été réalisées avec succès :
    - la substitution du gène lab muté chez la Drosophile (mutation qui entraîne la formation d'une tête très petite) par le gène Hox B1 de Poulet restaure la morphogenèse normale de la tête de Mouche ;
    - l'insertion du gène Hox B6 de Souris chez la Drosophile entraîne la transformation des antennes en pattes thoraciques ;
    - d'une manière générale, un gène homéotique de Vertébré possède son homologue chez les Insectes ; inséré par transgenèse interspécifique dans le génome de la Drosophile, le gène homéotique de la Souris ou de l'Homme induit des transformations homéotiques semblables à celles produites par son homologue ;
- la comparaison de tous les gènes homéotiques présents chez les Arthropodes et les Vertébrés met en évidence une grande similitude au niveau de l'homéoboîte. Cette similitude de séquence permet de dire que tous ces gènes sont homologues, c'est-à-dire qu'ils ont une origine commune ; ils forment donc une famille multigénique. Tous ces gènes se sont formés à partir d'un même gène ancestral par duplications géniques suivies de mutations.

	Domaine-B4 Homme	**Domaine-B7 Homme**	**Domaine-Dfd Drosophile**	**Domaine-Ubx Drosophile**
Domaine-B4 Homme	100			
Domaine-B7 Homme	83,3	100		
Domaine-Dfd Drosophile	88,3	83,3	100	
Domaine-Ubx Drosophile	75	90	73,3	100

Tableaux des identités (en %) entre plusieurs homéodomaines de gènes homéotiques de Vertébrés et d'Arthropodes (valeurs obtenues à l'aide d'un alignement réalisé avec le logiciel Anagène*)*

Le fait qu'il y ait plus de similitudes entre les gènes Hoxb4 de l'Homme et Dfd de la Drosophile qu'entre les gènes HoxB4 et HoxB7 indique que la duplication à l'origine de B4 et B7 est antérieure à la séparation des lignées des Insectes et des Vertébrés. Autrement dit, l'ancêtre commun à l'Homme et à la Drosophile avait un gène à l'origine du Dfd de la Drosophile et du B4 de l'Homme. En outre, il possédait un gène à l'origine du gène HoxB7 et du gène Ubx (plus forte similitude entre Ubx et B7 qu'entre HoxB4 et HoxB7). Le même raisonnement peut être tenu pour les gènes HoxB1, HoxB9, AbdB et Lab.

Des comparaisons du même type entre les gènes homéotiques des Vertébrés et de la Drosophile ont conduit à proposer le schéma évolutif suivant :



Ce schéma indique que l'ancêtre commun aux Vertébrés et aux Insectes possédait un complexe homéotique formé de six gènes, eux-mêmes résultant de duplications antérieures. Ces six gènes sont à l'origine des complexes Hom de la Drosophile et Hox des Vertébrés. L'Amphioxus est un Cordé qui ne possède qu'un seul complexe Hox. Les deux duplications à l'origine des quatre complexes existant chez tous les Vertébrés ont donc eu lieu dans la lignée des Cordés aboutissant à l'ancêtre commun à tous les Vertébrés.

Il est possible de faire travailler les élèves sur le schéma évolutif ci-dessus, les amener à extraire les informations essentielles puis à tester la validité de ce schéma à partir des séquences des gènes homéotiques de la Drosophile et de la Souris.

Maintien des innovations génétiques – *Exemple des globines*

Informations scientifiques

Cf. page 124 (classe terminale)

Pistes d'exploitation pédagogique

Les séquences de globines ont déjà été utilisées pour établir des relations de parenté entre les Vertébrés. Les élèves savent donc que tous les gènes bêta des globines (de même que les gènes alpha) proviennent d'un gène possédé par l'ancêtre commun à tous ces Vertébrés. Les différences entre les gènes des espèces actuelles résultent de mutations intervenues au cours de l'histoire des diverses lignées, **mutations qui ont réussi à se propager et à se fixer dans les populations successives**. La comparaison des séquences des gènes (ou des protéines) permet de rechercher si ces mutations sont réparties de façon quelconque ou non et, à partir des informations obtenues, de **réfléchir sur le devenir des mutations**.

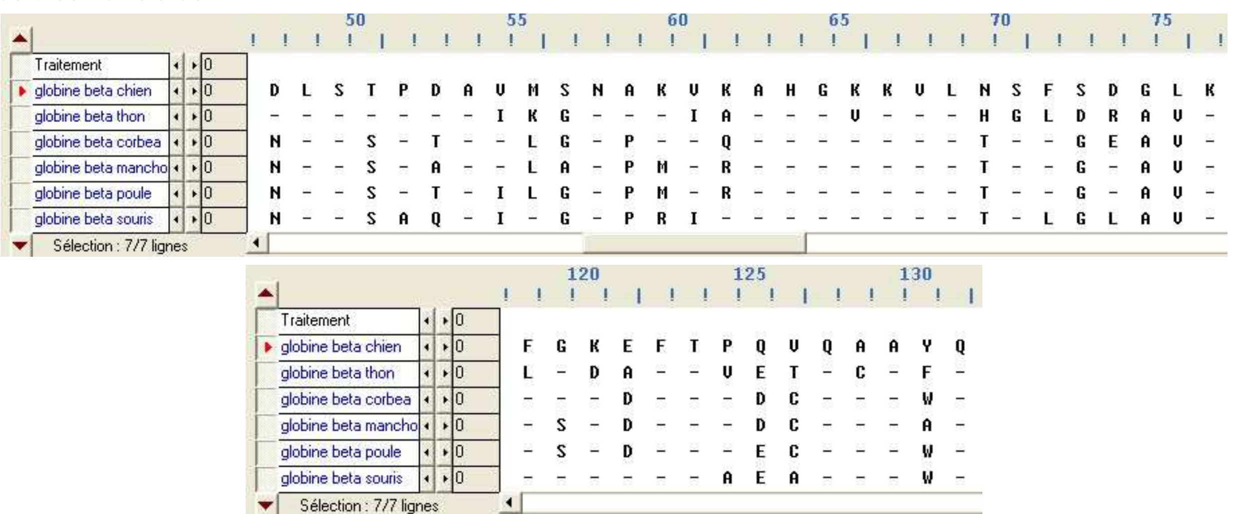
Séquences et documents

Fichiers des séquences

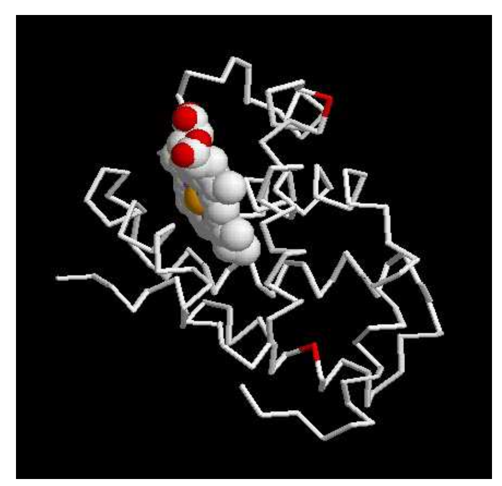
Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Maintien des innovations génétiques/Exemple des globines** permet d'atteindre **Gènes des globines** qui charge le fichier 6-globines-beta-Vertebres.edi affichant des séquences nucléiques de gènes de globines.

La comparaison de ces séquences permet de mettre en évidence des sites très variables et des sites conservés.

Sites très variables



Repérage de quelques sites très variables dans les globines bêta des Vertébrés. À certaines positions, on observe une grande variabilité des acides aminés selon les globines : c'est le cas en position 57 par exemple, ou encore en position 130



Localisation de ces sites sur une molécule 3D de bêta globine (en rouge, les sites particulièrement variables)

Sites conservés

Comparaison simple

					25					30					35					40			
Traitement	0																						
globine beta thon	0	D	I	G	P	K	A	L	A	R	C	L	I	V	Y	P	W	T	Q	R	Y	F	G
globine beta chien	0	E	V	-	G	E	-	-	G	-	L	-	-	-	-	-	-	-	-	-	F	-	D
globine beta corbea	0	E	C	-	G	E	-	-	-	-	L	-	-	-	-	-	-	-	-	-	F	-	A
globine beta mancho	0	E	C	-	A	E	-	-	-	-	L	-	-	-	-	-	-	-	-	-	F	-	A
globine beta poule	0	E	C	-	A	E	-	-	-	-	L	-	-	-	-	-	-	-	-	-	F	-	A
globine beta souris	0	K	V	-	G	E	T	-	G	-	L	-	-	-	-	-	-	-	-	-	F	-	D

Sélection : 0/7 lignes

		60					65					70
Traitement	0											
globine beta thon	0	I	A	A	H	G	V	K	V	L	H	G
globine beta chien	0	V	K	-	-	-	K	-	-	-	N	S
globine beta corbea	0	V	Q	-	-	-	K	-	-	-	T	S
globine beta mancho	0	V	R	-	-	-	K	-	-	-	T	S
globine beta poule	0	V	R	-	-	-	K	-	-	-	T	S
globine beta souris	0	-	K	-	-	-	K	-	-	-	T	S

Sélection : 0/7 lignes

							90					95					100					105				
Traitement	0																									
globine beta thon	0	Y	S	E	L	S	V	L	H	S	D	K	L	H	V	D	P	D	N	F	R	I	L	G	D	C
globine beta chien	0	F	A	K	-	-	E	-	-	C	-	-	-	-	-	-	-	E	-	-	K	L	-	-	N	V
globine beta corbea	0	F	A	Q	-	-	E	-	-	C	-	-	-	-	-	-	-	E	-	-	-	L	-	-	-	I
globine beta mancho	0	F	A	Q	-	-	E	-	-	C	-	-	-	-	-	-	-	E	-	-	-	L	-	-	-	I
globine beta poule	0	F	-	Q	-	-	E	-	-	C	-	-	-	-	-	-	-	E	-	-	-	L	-	-	-	I
globine beta souris	0	F	A	H	-	-	E	-	-	C	-	-	-	-	-	-	-	E	-	-	K	L	-	-	N	M

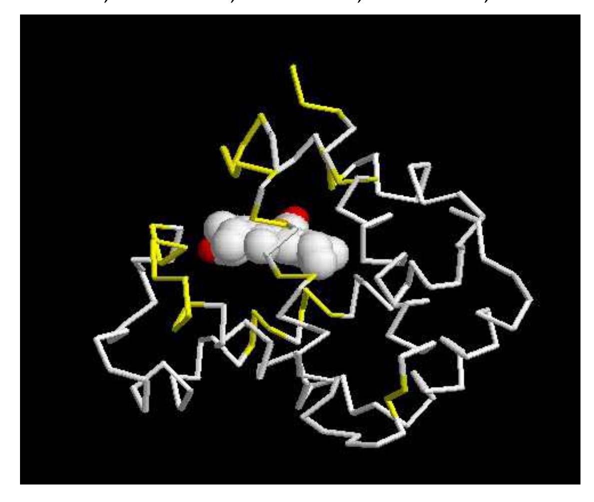
Sélection : 0/7 lignes

		115					120					125					130					135					140					145	
Traitement	0																																
globine beta thon	0	A	A	N	L	G	D	A	F	T	V	E	T	Q	C	A	F	Q	K	F	L	A	V	V	V	F	A	L	G	R	K	Y	H
globine beta chien	0	-	H	H	F	-	K	E	-	-	P	Q	V	-	A	-	Y	-	-	V	V	-	G	-	A	N	-	-	A	H	-	-	-
globine beta corbea	0	-	-	H	F	-	K	D	-	-	P	D	C	-	A	-	W	-	-	L	V	R	-	-	A	H	-	-	A	-	-	-	-
globine beta mancho	0	-	-	H	F	S	K	D	-	-	P	D	C	-	A	-	A	-	-	L	V	R	-	-	A	H	-	-	A	-	-	-	-
globine beta poule	0	-	-	H	F	S	K	D	-	-	P	-	C	-	A	-	W	-	-	L	V	R	-	-	A	H	-	-	A	-	-	-	-
globine beta souris	0	S	S	Y	F	-	K	E	-	-	A	-	A	-	A	-	W	-	-	L	V	V	G	-	A	T	-	-	S	H	-	-	-

Sélection : 0/7 lignes

Résultat d'une comparaison simple de globines bêta de quelques Vertébrés

On peut donc noter des sites « très conservés », c'est-à-dire n'ayant fixé aucune mutation : 32 à 40 ; 42 ; 62 à 64 ; 66 à 68 ; 88 à 89 ; 91 et 92 ; 94 à 100 ; 102 à 103 ; 106 à 107 ; 122 à 123 ; 140 à 141 ; 144 à 146.



Mise en évidence de ces sites conservés sur une molécule de globine bêta (en jaune : les sites conservés)

Les mutations apparaissent au hasard, avec *a priori* la même fréquence tout au long de la molécule. On constate cependant que certains sites sont particulièrement variables d'un taxon à l'autre, alors que d'autres sites sont très identiques.

La localisation sur la molécule 3D montre que les sites conservés sont **en règle générale situés autour de l'hème**. On peut donc supposer qu'un changement dans la séquence des acides aminés dans ces secteurs doit avoir des répercussions sur la fixation de l'hème ou la structure spatiale de la molécule dans cette région, et donc altérer la fonction de la globine. Ainsi, les individus porteurs de ces mutations doivent être « désavantagés », donc ces mutations se transmettent beaucoup moins ou pas du tout et finissent par être éliminées.

En revanche, des changements d'acides aminés qui ne sont pas importants pour le maintien de la structure spatiale de la globine ou la fixation de l'hème n'ont pas d'effet sur la survie ou la reproduction des individus qui les portent. Ils ne confèrent donc aucun avantage sélectif. Et pourtant, leur présence dans les populations des espèces actuelles indique que ces mutations apparues à différents moments de l'histoire des lignées se sont répandues dans les populations et transmises aux espèces successives des lignées. **La conservation de ces mutations dites neutres est aléatoire** (mécanisme de la dérive génique, hors programme).

Maintien des innovations génétiques – *Exemple des estérases – Résistance des Moustiques aux insecticides*

Informations scientifiques

À partir des années 1960, des insecticides organophosphorés (OP) ont été utilisés pour lutter contre les Moustiques, notamment dans la région de Montpellier. Parallèlement à cette utilisation, on a constaté l'apparition dans les régions traitées, et avec une fréquence croissante au cours du temps, de Moustiques (*Culex pipiens*) résistants à ces insecticides.

Des innovations génétiques sont à l'origine de cette résistance ; ces innovations sont apparues au hasard, mais une sélection positive des individus qui en étaient porteurs a dû avoir lieu, expliquant l'augmentation de fréquence des phénotypes résistants qui a été constatée au cours du temps.

Le mode d'action des insecticides organophosphorés OP

Les insecticides OP et les carbamates ont pour cible l'enzyme acétylcholine estérase (ACE, présente au niveau des synapses cholinergiques et ayant pour rôle l'hydrolyse du neuromédiateur acétylcholine) qu'ils inhibent. Les anomalies de fonctionnement du système nerveux qui en résultent entraînent la mort de l'Insecte sensible.

Cependant, pour qu'un insecticide soit efficace il faut qu'il puisse atteindre sa cible, c'est-à-dire ici l'espace synaptique. Il doit donc pénétrer dans l'organisme de l'Insecte et circuler dans son milieu intérieur.

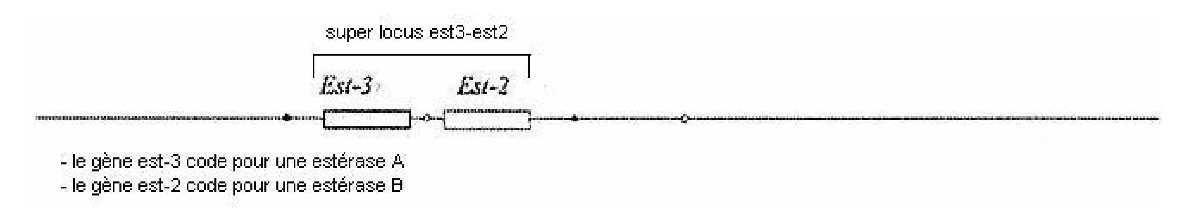
Les Moustiques non résistants peuvent lutter naturellement contre de faibles doses d'insecticides OP

La résistance aux insecticides OP et aux carbamates peut s'expliquer par des mécanismes qui empêchent ces insecticides d'atteindre leur cible, ou par une sensibilité moindre de la cible (ACE).

Les estérases, enzymes produites naturellement par les Moustiques, hydrolysent les liaisons ester, notamment celles des molécules des insecticides OP et des carbamates (comme le propoxur). Il existe chez le Moustique deux sortes d'estérases appelées A et B, qui hydrolysent toutes les deux des liaisons ester, mais avec une spécificité relative à la position des liaisons. Les gènes codant pour les estérases ne s'expriment pas dans tous les tissus de l'Insecte ; cette expression a surtout lieu dans la paroi du tube digestif, la zone cellulaire cutanée sous-hypodermique et les ganglions cérébraux et thoraciques, les deux premiers tissus constituant les principales voies de pénétration de l'insecticide dans l'organisme, les derniers contenant la cible des insecticides. Les estérases permettent donc d'empêcher l'insecticide de pénétrer et de circuler dans l'organisme de l'Insecte, donc d'atteindre sa cible. Il s'agit ici d'un mécanisme de **détoxication par métabolisation de l'insecticide.** Le rôle physiologique des estérases est inconnu actuellement…

Déterminisme génétique de la production d'estérases chez les Moustiques non résistants aux insecticides OP

Ces estérases sont codées par deux gènes (Est 3 codant pour l'estérase A et Est 2 codant pour l'estérase B) dont les loci sont situés sur le même chromosome et espacés de 2 à 6 kb. Il existe un polymorphisme important pour chacun de ces gènes. Ce polymorphisme est neutre, c'est-à-dire sans conséquences sur l'activité enzymatique.



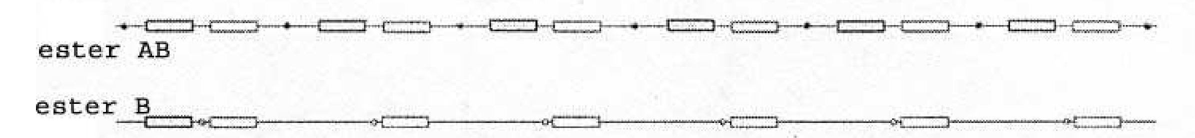
Dans un souci de simplification, ces gènes seront désignés par les lettres A et B dans la banque de données fournies, et les allèles de chacun de ces gènes sont désignés par des numéros (A1, A2… B2, B4…)

La résistance des Moustiques aux insecticides OP

La production d'estérases par les Moustiques sensibles est très faible, insuffisante pour empêcher l'insecticide d'agir. Chez certaines souches résistantes, la production d'estérases en beaucoup plus grande quantité est le mécanisme à l'origine de la résistance. Il ne s'agit donc pas d'un changement qualitatif (efficacité des estérases) mais d'un changement quantitatif (l'augmentation de la quantité d'estérases produites peut être considérable, jusqu'à représenter 6 à 10 % des protéines totales de l'Insecte). Un test au papier-filtre permet d'évaluer la quantité d'estérases produites par un Moustique.

Le génome des Moustiques résistants par production accrue d'estérases

Les Moustiques dont la résistance est due à une production accrue d'estérases ont un génome caractérisé par une **amplification des gènes Est3 (ou gène A) et Est2 (ou gène B)** : chacun de ces gènes est présent en de nombreux exemplaires. L'innovation génétique consiste donc en **duplications géniques**. Tous les gènes résultant de ces duplications sont identiques et s'expriment, ce qui explique la production accrue d'estérases. On désigne par « ester » le « supergène » résultant de l'amplification (ester B pour le résultat de l'amplification du gène B, ester A-B pour le résultat de l'amplification de l'ensemble des deux gènes A et B).



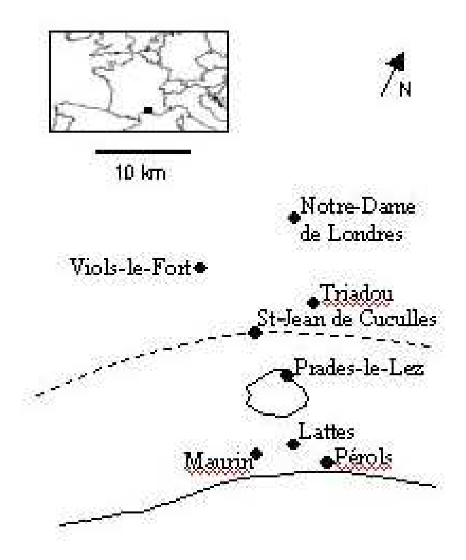
Selon l'allèle du gène qui est amplifié, on distingue plusieurs « supergènes ester » : ester B1 (plusieurs copies de l'allèle B1), ester A2-B2 (plusieurs copies de l'ensemble des deux allèles A2-B2)…

L'existence de plusieurs « supergènes ester » différents (ester B2, ester B4, ester A2-B2, ester A4-B4…) indique que **plusieurs innovations génétiques à l'origine de la résistance sont intervenues de façon indépendante dans les populations de Moustiques** (duplications à partir de l'allèle B2, duplication à partir de l'allèle B4…). Cependant, le polymorphisme des gènes A et B est très nettement supérieur à celui des « supergènes ester » ; cela indique donc que les mutations à l'origine des « supergènes ester » sont relativement rares. D'autre part, on a pu constater que des allèles ester apparus en premier dans une région se retrouvaient ensuite dans une autre région assez éloignée ; cela s'interprète comme le résultat de migrations des Moustiques (par transport aérien ou maritime par exemple) et non comme le résultat de deux mutations identiques indépendantes. Le niveau d'amplification varie selon les allèles ester : il peut atteindre cent copies du gène ou plus pour ester B1, tandis qu'il ne dépasse pas quelques copies pour ester A4-B4. De plus, pour un même allèle (par exemple, ester B1) le niveau d'amplification peut être variable selon les individus au sein d'une population sauvage ou entre des populations différentes.

Enfin, il existe un allèle ester A1 pour lequel la production accrue d'estérases n'est pas due à l'amplification du gène A1, mais à un changement dans la régulation de l'expression de ce gène. Comme les processus de régulation génique ne sont pas au programme de terminale S, nous n'avons pas retenu cet allèle dans la banque de données.

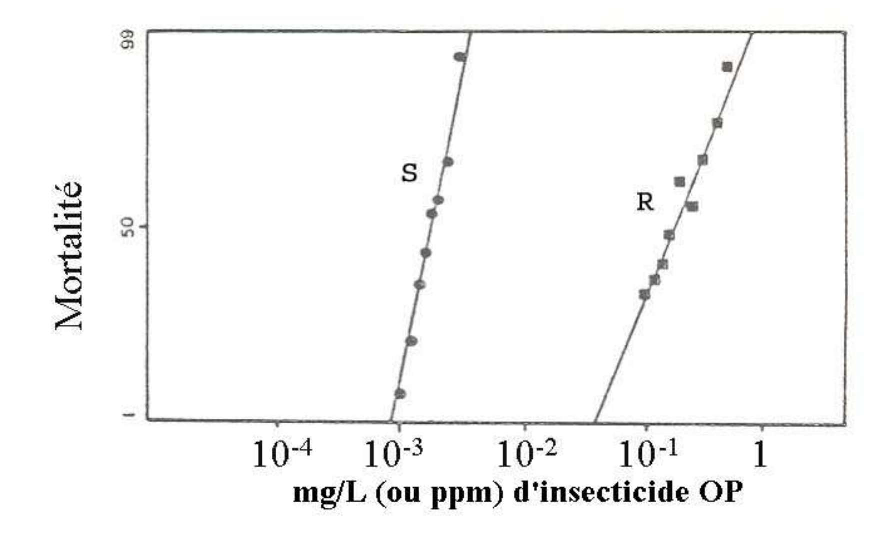
Les observations de terrain et les données expérimentales relatives à l'action de l'environnement sur la fréquence des phénotypes résistants

Des chercheurs ont déterminé le pourcentage de Moustiques résistants aux insecticides OP dans la région de Montpellier, dans une zone où il y a eu épandage d'insecticides depuis les années 1970, et dans une zone non soumise aux insecticides.



Les sites d'observation et de mesures.
La ligne pointillée indique la séparation entre la zone non traitée par les insecticides OP (au nord) et la zone traitée (près des côtes)

Pour ce faire, ils prélèvent des larves qu'ils placent dans des solutions d'insecticides de concentration croissante. Ils déterminent le nombre de larves mortes au bout de vingt-quatre heures, ce qui leur permet d'établir une courbe de mortalité.



Courbes de mortalité expérimentale des larves en fonction des doses d'insecticides
S : souche sensible, R : souche résistante Les résultats sont présentés sous forme de graphique, les doses utilisées nécessitant l'utilisation d'une échelle logarithmique en abscisse

En opérant ainsi dans la région de Montpellier, ils ont obtenu les résultats indiqués dans le tableau et le graphe ci-dessous :

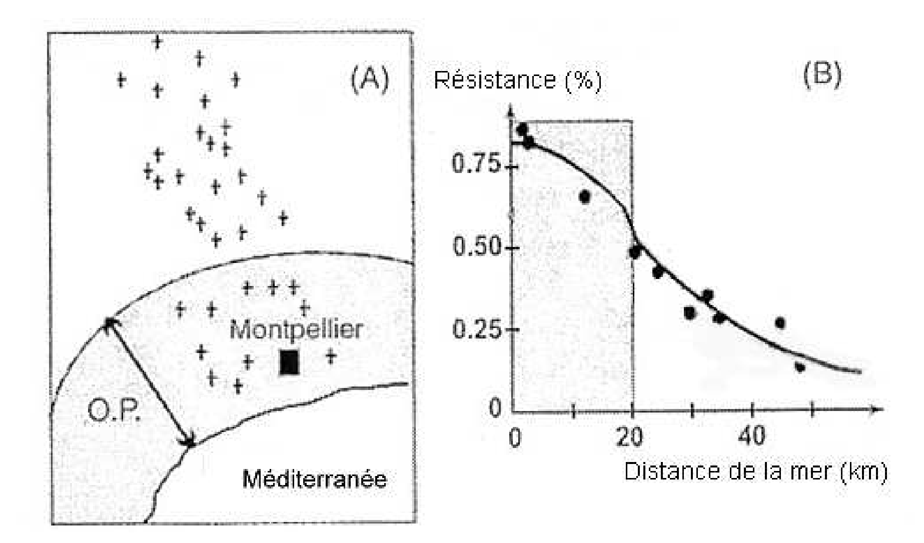
Communes	Morts	Vivants	Total	Distance	%vivants
Pérols	21	147	168	2,2	0,88
Maurin	0	227	227	5	1,00
Lattes	42	95	137	5,2	0,69
Prades-le-Lez	20	112	132	18	0,85
Triadou	82	149	231	24	0,65
St-Jean de Cuculle	73	81	154	26	0,53
Viols-le-Fort	93	74	167	30	0,44
Notre-Dame de Londres	155	41	196	35	0,21

Eté 2002
Fréquence des survivants
1,00 0,90 0,80 0,70 0,60 0,50 0,40 0,30 0,20 0,10 0,00
0 10 20 30 40
(Km depuis la mer)

Les larves, échantillonnées le long du cline (nom de la commune indiquée, ainsi que la distance à la mer), ont été soumises à une dose d'insecticide qui tuait tous les Moustiques avant les années 1968. Les larves mortes et vivantes ont été comptées au bout de 24 heures

On constate que la fréquence des phénotypes résistants est très forte dans la région soumise aux insecticides et qu'elle diminue au fur et à mesure que l'on s'éloigne de cette zone.

Des chercheurs ont déterminé le génotype des Moustiques recueillis dans la région de Montpellier en ce qui concerne les gènes des estérases. À partir des données obtenues, ils ont calculé la fréquence des allèles de résistance ester en fonction de la distance au bord de mer.



On peut alors constater que la fréquence des allèles de résistance est nettement supérieure dans la zone traitée et qu'elle diminue régulièrement lorsque l'on s'éloigne de cette zone. La recherche d'une explication à ces observations conduit naturellement à faire intervenir la sélection positive des Moustiques porteurs d'allèles de résistance dans la zone traitée…

Une étude expérimentale a été menée pendant quatre ans en laboratoire afin de tester l'influence de facteurs du milieu de vie sur les variations de fréquence alléliques et phénotypiques.

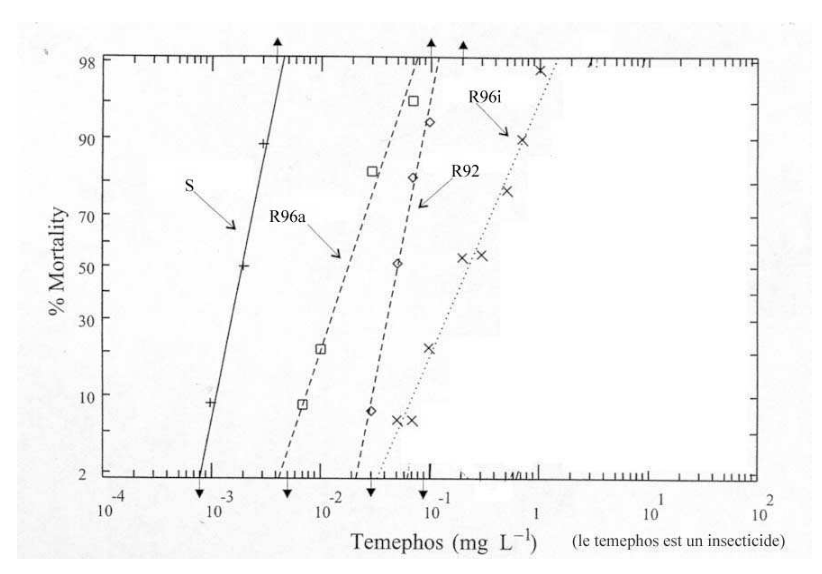
La souche « Edit » (R92 sur le graphe) est une souche de Moustiques résistants obtenue par les chercheurs en 1992 à la suite de nombreux croisements. La résistance de la souche est due à l'allèle EsterB1 qui entraîne une production considérablement amplifiée de l'estérase B1. La souche est initialement homozygote pour l'allèle B1 : tous les Moustiques ont deux allèles B1 avec le même nombre de répétitions de B1.

Les chercheurs ont divisé cette population de Moustiques de lignée pure pour EsterB1 en deux. La première fraction a été élevée pendant quatre ans dans un milieu sans insecticide (R96a). La seconde fraction a été soumise, à chaque génération, au dernier stade larvaire, à l'action d'un insecticide OP à une concentration entraînant une mortalité d'au moins 80 % des larves (souche « Sedit » – R96i). L'expérimentation permet donc de suivre l'évolution du phénotype des Moustiques initialement de même génotype, en fonction de l'environnement.

On met alors en évidence deux phénotypes macroscopiques au sein des populations de Moustiques : le phénotype « sensible aux insecticides OP » et le phénotype « résistant aux insecticides OP ».

Les scientifiques utilisent la DL50 pour traduire la résistance des Moustiques. Cette valeur correspond à la concentration d'insecticides pour laquelle on obtient 50 % de mortalité des larves.

En 1996, on a déterminé les courbes de mortalité des deux populations Edit96 (R96a) et SEdit (R96i) et on les a comparées à celle de Edit92 (R92).



Graphique présentant les résultats d'une expérimentation menée sur quatre ans pour étudier l'influence d'un facteur de l'environnement (les insecticides OP) sur la fréquence des phénotypes des Moustiques

La courbe de mortalité de « Sedit » (R96i) est décalée vers la droite par rapport à celle de « Edit92 » (R92). Cela indique que globalement, la population « Sedit » (R92i) est plus résistante que la population « Edit » (R92) dont elle est issue. La dose DL50 pour la souche « Edit92 » (R92) est environ 5.10^{-2} et pour « SEdit96 » (R96i) elle est environ de $2{,}5.10^{-1}$, ce qui donne un rapport de 5 environ.

La courbe de « Edit96 » (R96a) est décalée vers la gauche par rapport à « Edit92 » (R92) : globalement, la population « Edit96 » (R96a) est moins résistante que celle de « Edit92 » (R92). La DL50 de « Edit96 » (R96a) est de 2.10^{-2}, ce qui donne un rapport vis-à-vis de « Edit92 » (R92) de 0,4.

On constate donc une sélection positive des Moustiques résistants dans la population soumise au fil des générations à l'insecticide, mais aussi une sélection en faveur d'organismes moins résistants dans la population élevée en absence d'insecticide. Cela semble indiquer un désavantage des Moustiques résistants dans le milieu sans insecticide. Une première explication de ce coût de la résistance, certes sans support expérimental, consiste à dire qu'en l'absence d'insecticide, la quantité considérable d'estérase produite l'est pour rien. Il est possible que la production d'une telle quantité de protéines, comparativement aux Moustiques sensibles, affecte d'une certaine façon la valeur sélective. On a également montré que les souches produisant beaucoup d'estérases étaient beaucoup plus sensibles aux bactéries Wolbachia, ces dernières ayant une action néfaste sur la reproduction des Moustiques.

Les recherches ont été poursuivies sur les souches « Edit » (R92) et « Sedit » (R96i) en comparant l'évolution du génotype des deux populations. Chez quelques individus de chaque population, on a déterminé le nombre de répétitions de B1 dans l'allèle EsterB1 qu'ils possédaient. Le niveau d'amplification dans les deux populations « Edit96 » (R96a) et « SEdit96 » (R96i) était bien différent : une moyenne de 23 +/- 7 copies chez « Sedit » (R96i), et de 11+/- 2 chez « Edit » (R92). Cela indique que dans l'histoire évolutive des deux populations de nouvelles mutations conduisent à un polymorphisme génique avec une sélection positive des individus porteurs d'un grand nombre de répétitions en milieu avec insecticide et de ceux dont le niveau d'amplification est faible dans le milieu sans insecticide. Ainsi, il y a une interaction permanente entre innovations génétiques et environnement.

NB : Le plus difficile est bien sûr de faire saisir que le milieu n'oriente pas la nature des mutations, mais que celles-ci sont sélectionnées en fonction des caractéristiques de celui-ci.

Dans un article de *Médecine Sciences* de décembre 2003, Michel Raymond *et al*. écrivent : « Il est important de comprendre que le Moustique ne mute pas pour résister aux insecticides ! De très nombreuses mutations préexistent dans les immenses populations de Moustiques. Lorsque des insecticides sont présents dans l'environnement, les Moustiques qui ont des mutations favorables à leur survie se reproduisent alors que les Moustiques sensibles aux toxiques meurent ! Les mutations conférant aux Moustiques la capacité de résister aux OP ne sont pas engendrées mais sélectionnées par le milieu. »

Pistes d'exploitation pédagogique

Les données fournies dans les thèmes d'étude sont utiles pour bâtir des séquences pédagogiques permettant d'aborder les mécanismes de l'évolution, tant en ce qui concerne les innovations génétiques (duplication génique, suivie ou non de mutations aléatoires), que l'influence de certains facteurs du milieu de vie sur le maintien de ces innovations génétiques au sein d'une population.

Il s'agit de données expérimentales, d'observations sur le terrain et de séquences nucléiques, fournies par Michel Raymond et son équipe (Institut des sciences de l'évolution de l'université de Montpellier I).

Innovations génétiques : apparition de nouveaux allèles par mutations

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Maintien des innovations génétiques/Exemple des estérases – Résistance des Moustiques aux insecticides/Innovations génétiques** permet d'atteindre **Duplication** qui charge le fichier duplic-esterases.edi affichant en tête de liste les séquences :

- AllèleB1complet : séquence nucléique de la partie strictement codante du gène B allèle B1 (du codon d'initiation au codon stop) – le gène B, ou Est2, code pour une estérase B ;

- AllèleB2complet : séquence nucléique de la partie strictement codante du gène B allèle B2 (du codon d'initiation au codon stop) – le gène B, ou Est2, code pour une estérase B.

Comparaison des allèles B1 et B2

De nouveaux allèles d'un même gène apparaissent par mutations. Les allèles B1 et B2 diffèrent par cinq mutations dont une seule apparaît dans la séquence protéique - 39 (C/T) - 59 (A/G) - 63 (A/G) - 66 (G/T) - 67 (TC) ; seule la 59 a des conséquences : AAC/AGC donc Asn (N) au lieu de Ser (S). Les estérases codées par ces allèles sont fonctionnelles.

Les résultats de ces comparaisons sont donc utilisables pour aborder les notions de mutations neutres et de mutations muettes.

Innovations génétiques : phénomène de duplication génique

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Maintien des innovations génétiques/Exemple des estérases – Résistance des Moustiques aux insecticides/Innovations génétiques** permet d'atteindre **Duplication** qui charge le fichier duplic-esterases.edi affichant entre autres les séquences :

- GèneBallèle1 : séquence simplifiée du gène B allèle B1 – afin de faciliter les traitements des séquences, on a limité cette séquence aux soixante-neuf premiers nucléotides du gène + codon stop ;

- B1R : séquence simplifiée du « locus B » d'un Moustique résistant de la souche B1R (appelée aussi EsterB1) – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux soixante-neuf premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence B1R représente donc huit gènes B, allèle B1, disposés à la suite les uns des autres, et séparés par des séquences non codantes représentées par 12N ;

- A2B2 : séquence simplifiée du « locus AB » – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux soixante-neuf premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence A2B2 représente donc un gène A, allèle A2, simplifié, suivi de 12N (séquence non codante) puis d'un gène B allèle B2, simplifié. ;

- A2B2R : séquence simplifiée du « locus AB » d'un Moustique résistant de la souche A2B2R (appelée aussi Ester A2B2) – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux soixante-neuf premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence A2B2R représente donc quatre ensembles A2B2, disposés à la suite les uns des autres, et séparés par des séquences non codantes représentées par 12N (une séquence de 12N séparant également les gènes A2 et B2 dans chaque ensemble A2B2).

Comparaison des séquences B1 et B1R, présentes respectivement chez les Moustiques sensibles et résistants

Chez le Moustique résistant (génome B1R), on constate l'existence de plusieurs exemplaires (huit dans les séquences fournies) du gène B1 disposés « en tandem », tous identiques entre eux, ce qui peut s'interpréter par un phénomène de duplication génique du gène B1 (non encore suivi de mutation).

Comparaison des séquences A2B2 et A2B2R, présentes respectivement chez les Moustiques sensibles et résistants

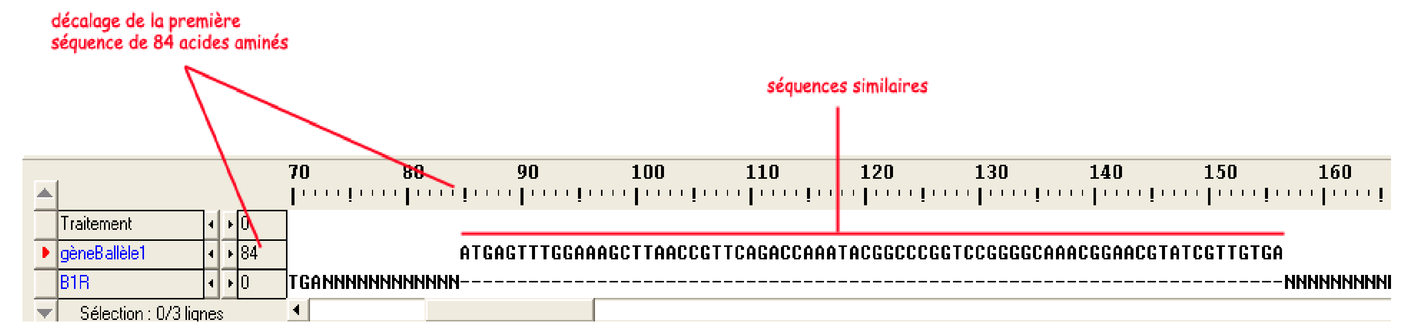
Chez le Moustique résistant (génome A2B2R), on constate l'existence de plusieurs exemplaires (quatre dans les séquences fournies) de l'ensemble « A2B2 », tous identiques entre eux, ce qui peut s'interpréter par un phénomène de duplication génique de l'ensemble A2B2.

Remarques

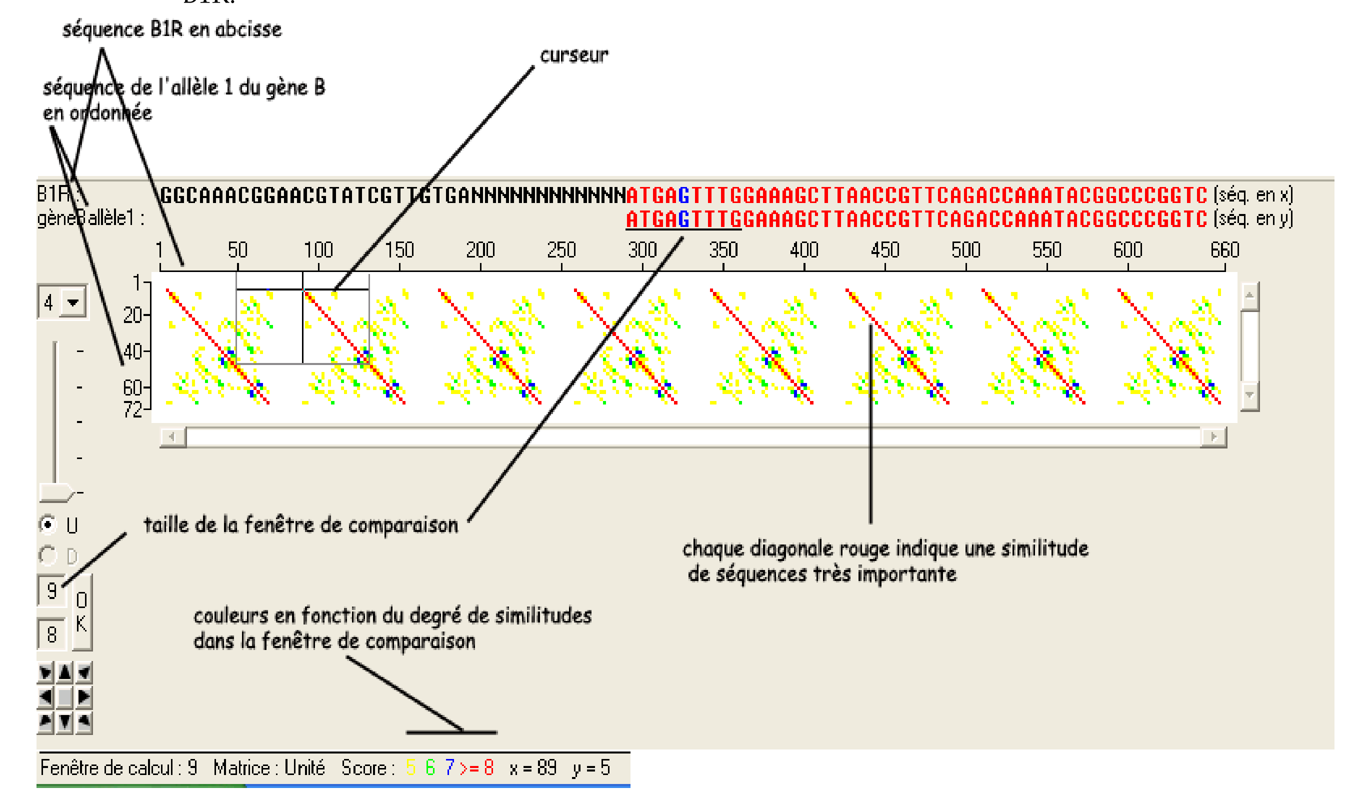
Il existe deux façons de comparer les séquences proposées avec le logiciel *Anagène* :

- utilisation de la fonctionnalité **Comparaison simple de séquences** :

  Pour comparer des séquences de longueur différente et mettre en évidence la répétition du gène à l'identique, il faut pouvoir décaler la séquence la plus courte de façon à la comparer à différentes zones de la séquence la plus longue ; on peut procéder de la façon suivante :
  - o demander une comparaison simple en sélectionnant en premier le génome le plus court ;
  - o repérer dans le génome formé de plusieurs gènes la séquence de NNNN qui sépare deux gènes et noter la position à laquelle se finit cette séquence non codante (exemple : 84) ;
  - o inscrire en face du nom de la première séquence, à la place du 0, le numéro relevé précédemment (84) et valider.



- utilisation de la fonctionnalité **DotPlot** :
  - o sélectionner les deux séquences à comparer et demander le Dot Plot ;
  - o l'affichage permet de constater que la séquence B allèle 1 se retrouve à l'identique huit fois dans B1R.



Innovations génétiques : phénomène de duplication génique suivie de mutations

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Maintien des innovations génétiques/Exemple des estérases – Résistance des Moustiques aux insecticides/Innovations génétiques** permet d'atteindre **Duplication** qui charge le fichier duplic-esterases.edi affichant entre autres les séquences :

- AllèleA2complet : séquence nucléique de la partie strictement codante du gène A allèle A2 (du codon d'initiation au codon stop) – le gène A, ou Est3, code pour une estérase A ;

- GèneAallèle2 : séquence simplifiée du gène A allèle A2 – afin de faciliter les traitements des séquences, on a limité cette séquence aux soixante-neuf premiers nucléotides du gène + codon stop ;

- GèneBallèle2 : séquence simplifiée du gène B allèle B2 – afin de faciliter les traitements des séquences, on a limité cette séquence aux soixante-neuf premiers nucléotides du gène + codon stop.

Comparaison des gènes A (allèle A2) et B (allèle B2)

Les similitudes importantes suggèrent une origine commune des deux gènes (on peut alors évoquer une duplication génique), mais cette duplication a dû être suivie de mutations, ce qui a abouti à deux gènes aujourd'hui différents (qui codent tous deux pour des estérases, mais aux propriétés légèrement différentes).

Influence d'un facteur du milieu de vie sur le maintien des innovations génétiques

Séquences et documents

Séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Maintien des innovations génétiques/Exemple des estérases – Résistance des Moustiques aux insecticides/Sélection naturelle** permet d'atteindre **Sélection** qui charge le fichier select-moustiques.edi affichant les séquences :

- gène B allèle 1 : séquence simplifiée du gène B allèle B1 – afin de faciliter les traitements des séquences, on a limité cette séquence aux soixante-neuf premiers nucléotides du gène + codon stop ;

- R92 : séquence simplifiée du « locus B » d'un Moustique résistant de la souche R92 utilisée par les chercheurs pour l'expérimentation – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux soixante-neuf premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence R92 représente donc quatorze gènes B, allèle B1, disposés à la suite les uns des autres, et séparés par des séquences non codantes représentées par 12N ;

- R96a-1 : séquence simplifiée du « locus B » d'un Moustique résistant de la souche R96a (ou Edit96), obtenue à partir de la souche R92 élevée dans un environnement sans insecticides OP pendant quatre ans – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux soixante-neuf premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence R92 représente donc sept gènes B, allèle B1, disposés à la suite les uns des autres, et séparés par des séquences non codantes représentées par 12N ;

- R96a-2 : séquence simplifiée du « locus B » d'un Moustique résistant de la souche R96a (ou Edit96), obtenue à partir de la souche R92 élevée dans un environnement sans insecticides OP pendant quatre ans – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux soixante-neuf premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence R92 représente donc dix gènes B, allèle B1, disposés à la suite les uns des autres, et séparés par des séquences non codantes représentées par 12N ;

- R96a-3 : séquence simplifiée du « locus B » d'un Moustique résistant de la souche R96a (ou Edit96), obtenue à partir de la souche R92 élevée dans un environnement sans insecticides OP pendant quatre ans – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux soixante-neuf premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence R92 représente donc treize gènes B, allèle B1, disposés à la suite les uns des autres, et séparés par des séquences non codantes représentées par 12N ;

- R96i-1 : séquence simplifiée du « locus B » d'un Moustique résistant de la souche R96i (ou SEdit96), obtenue à partir de la souche R92 soumise aux insecticides OP pendant quatre ans – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux soixante-neuf premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence R92 représente donc dix-huit gènes B, allèle B1, disposés à la suite les uns des autres, et séparés par des séquences non codantes représentées par 12N ;

- R96i-2 : séquence simplifiée du « locus B » d'un Moustique résistant de la souche R96i (ou SEdit96), obtenue à partir de la souche R92 soumise aux insecticides OP pendant quatre ans – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux soixante-neuf premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence R92 représente donc vingt-deux gènes B, allèle B1, disposés à la suite les uns des autres, et séparés par des séquences non codantes représentées par 12N ;

- R96i-3 : séquence simplifiée du « locus B » d'un Moustique résistant de la souche R96i (ou Sedit96), obtenue à partir de la souche R92 soumise aux insecticides OP pendant quatre ans – afin de faciliter les traitements des séquences, on a limité la séquence de chaque gène aux 69 premiers nucléotides + codon stop ; les gènes sont séparés par 12N, correspondant à des séquences non codantes – cette séquence R92 représente donc vingt-trois gènes B, allèle B1, disposés à la suite les uns des autres, et séparés par des séquences non codantes représentées par 12N.

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Maintien des innovations génétiques : exemple de la résistance des Moustiques/Un modèle : la résistance aux insecticides chez les Moustiques/Innovations génétiques** et/ **Sélection naturelle** permettent d'atteindre :

- **Courbe de mortalité** qui charge le fichier MortaliteSR.jpg : graphique de résultats expérimentaux (échelle logarithmique). On soumet des larves de Moustiques (au dernier stade larvaire) à des doses croissantes d'insecticides et on évalue le pourcentage de mortalité en fonction des concentrations utilisées ;

- **Production d'estérases** qui charge le fichier quantiteesterase.jpg : présentation de résultats expérimentaux d'un test au papier-filtre permettant d'évaluer les quantités d'estérase produites par des Moustiques ;

- **Répartition des phénotypes** qui charge le fichier repphenomoustiques.jpg : graphique de répartition des phénotypes de Moustiques en fonction de la distance à la côte, dans la région de Montpellier. Ce graphique a été établi à partir de mesures réalisées en 1992 ;

- **Fréquence des génotypes** qui charge le fichier frequencegenotypes-esterases.jpg : graphique présentant la fréquence de quelques génotypes pour le gène des estérases chez les Moustiques, en fonction de la distance la côte.

- **Fréquence allèles résistants** qui charge le fichier resistance.jpg : graphique présentant la fréquence allèles « résistants » pour le gène des estérases chez les Moustiques, en fonction de la distance la côte ;

- **Étude expérimentale** qui charge le fichier etude-experimentale-moustiques.jpg : graphique présentant les résultats d'une expérimentation menée sur quatre ans pour étudier l'influence d'un facteur de l'environnement (les insecticides OP) sur la fréquence des phénotypes des Moustiques.

Comparaison du génome de la souche R92 (ou Edit 92) avec l'allèle B1

Le génome de la souche R92 comprend quatorze exemplaires du gène B1, tous ces exemplaires étant issus d'un allèle B1 par duplications successives.

Comparaison des génomes des souches R96a-1, R96a-2, R96a-3, R96i-1, R96i-2, R96i-3 avec l'allèle B1 (ou encore avec le génome de la souche R92)

Les souches placées pendant quatre ans dans un environnement riche en insecticides OP présentent un génome où le nombre d'exemplaires du gène B1 est supérieur à celui de la souche d'origine ; les souches placées pendant quatre ans dans un environnement sans insecticides OP présentent un génome où le nombre d'exemplaires du gène B1 est inférieur à celui de la souche d'origine. Ce constat permet de poser le problème de l'influence d'un facteur de l'environnement (les insecticides OP) sur le maintien des innovations génétiques (il faut bien sûr partir du principe que les innovations génétiques se font au hasard !).

Documents complémentaires

Évaluation de la quantité d'estérases produite par différentes souches de Moustiques

Il est possible d'évaluer expérimentalement la quantité d'estérases produites par un Moustique.

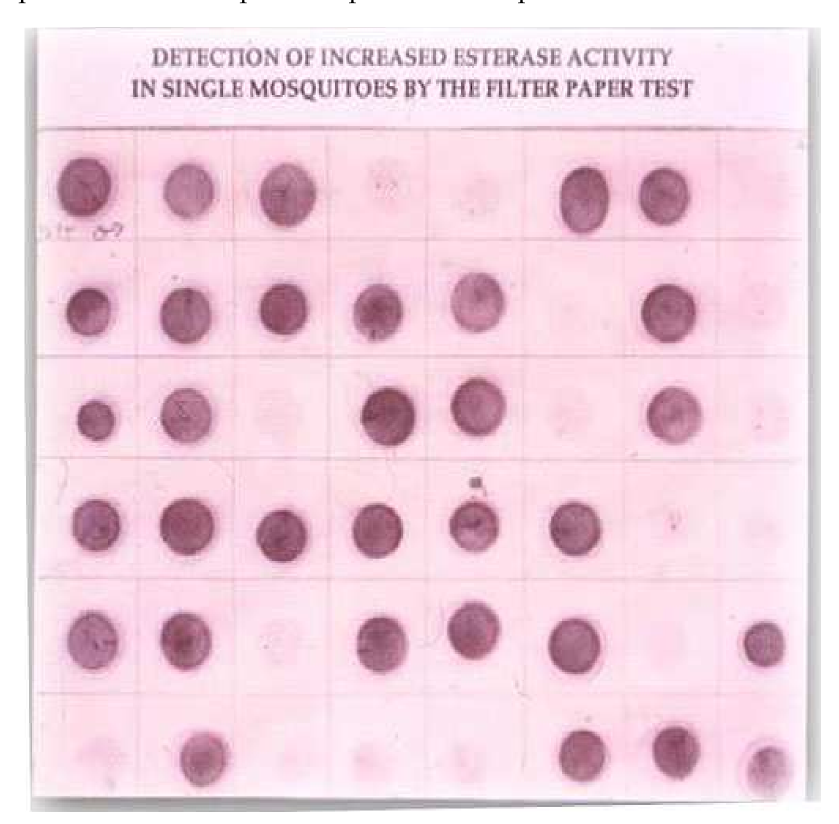
Protocole expérimental

On écrase des Moustiques recueillis sur un papier-filtre, puis on ajoute les réactifs suivants :

- un mélange de deux substrats sur lesquels l'estérase agit en les coupant ;
- un réactif qui colore en rouge l'un des produits obtenus.

Le substrat chromogène est un mélange de alpha-naphthyl acétate (alpha NA) et de bêta-naphthyl acétate (bêta NA). À la rigueur, un seul des deux peut être utilisé, mais habituellement, c'est un mélange des deux, dans un rapport 1/1, les estérases A utilisent préférentiellement alpha NA, et les estérases B agissent préférentiellement sur bêta NA. L'estérase coupe ces deux substrats, en libérant du alpha (ou bêta) naphthol, qui se colore en présence de Fast Garnet GBC (poudre à rajouter dans le tampon de coloration). Le produit plus Fast Garnet donne un précipité rouge, qui donne la coloration sur le papier-filtre ;

- une fois le Moustique écrasé sur le papier, la révélation se fait entre cinq et quinze minutes, si tous les réactifs sont prêts, chaque tâche correspondant à un Moustique écrasé, l'intensité de la coloration obtenue étant proportionnelle à la quantité d'estérase produite par ce Moustique.



Évaluation de la quantité d'estérases produite par différentes souches de Moustiques : résultats expérimentaux d'un test au papier-filtre

Ce protocole expérimental a été appliqué à une population de Moustiques de la région de Montpellier. On observe alors une variété phénotypique au sein de cette population de Moustiques, certains produisant beaucoup d'estérases, d'autres très peu.

Ces résultats sont à mettre en relation avec le document présentant la variabilité des populations de Moustiques au niveau du phénotype macroscopique (résistance ou sensibilité aux insecticides).

Maintien des innovations génétiques – *Exemple de la G6PD – Paludisme, sélection naturelle*

Informations scientifiques

Cf. page 138 (classe terminale)

Pistes d'exploitation pédagogiques

Les données sur les allèles de la G6PD ont permis d'établir que ceux-ci résultent de mutations à partir d'un gène ancestral G6PDB et de réfléchir ainsi à la filiation entre ces allèles. Ces allèles ont aussi une fréquence très variable suivant les populations. La recherche d'explications à ces différences permet de renforcer la notion de sélection naturelle positive.

Fréquence des allèles d'un gène et sélection naturelle

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Maintien des innovations génétiques/Exemple de la G6PD – Paludisme, sélection naturelle** permet d'atteindre **Allèles de la G6PD Sélection** qui charge le fichier G6PD-HS.edi affichant les séquences nucléiques strictement codantes de quelques allèles du gène de la G6PD.

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Maintien des innovations génétiques : paludisme, G6PD, sélection naturelle** permet d'atteindre :

- **Rôle G6PD** qui charge le fichier roleG6PD.bmp fournissant des informations scientifiques, sous forme d'un petit texte, concernant la G6PD et les conséquences d'une réduction de l'efficacité de l'enzyme de la G6PD ;
- **Fréquence allèles G6PD** qui charge le fichier FRQG6PD.bmp affichant la fréquence de quelques allèle de la G6PD ;
- **Carte répartition G6PD déficiente** qui charge le fichier reparti-G6PD.bmp affichant la carte de la déficience en G6PD à l'échelle mondiale : elle permet de repérer les régions du monde où les déficiences en G6PD sont relativement fréquentes, et de souligner la fréquence élevée en Afrique tropicale ;
- **Carte-paludisme-2002** qui charge le fichier palu-2002.bmp montrant la répartition du paludisme dans le monde en 2002, permet d'établir une première corrélation entre la répartition du paludisme et les zones où la G6PD déficience est répandue ;
- **Carte paludisme passé** qui charge le fichier palu-passe.bmp montrant la répartition du paludisme dans le monde par le passé, révèle que l'éradication du paludisme dans le bassin méditerranéen est récente et permet donc d'affiner la corrélation « paludisme/déficience ». Les discordances dans les Amériques s'expliquent par l'histoire des populations africaines ;
- **Localisation Vanuatu** qui charge le fichier local-vanuatu.bmp révélant la localisation géographique de l'archipel de Vanuatu ;
- **Fréquence allèles G6PD-Vanuatu** qui charge le fichier G6PD-Vanuatu.bmp affichant la fréquence des allèles de la G6PD dans les îles de l'archipel et permettant de retrouver la corrélation « paludisme/déficience » à l'échelle locale ;
- **Données épidémiologiques Gambie-Kenya** qui charge le fichier gambie-kenya.bmp affichant la fréquence des allèles G6PDA- dans deux populations d'enfants paludéens et non paludéens et qui montre que dans une même population la fréquence des allèles G6PDA- est statistiquement plus faible chez les enfants paludéens que chez les enfants non paludéens ;
- **Fréquence hématies parasitées** qui charge le fichier freq-hematies.bmp affichant la fréquence des hématies parasitées par le plasmodium chez les filles paludéennes et hétérozygotes G6PDA-/G6PDB ou G6PDA-/G6PDA. Elle montre le moindre développement du parasite dans les hématies G6PD déficientes que dans les hématies G6PD fonctionnelles ;
- **Étude expérimentale G6PD** qui charge le fichier experG6PD.bmp présentant l'étude *in vitro* de l'infection par le *Plasmodium* d'hématies G6PD déficientes et d'hématies G6PD fonctionnelles qui révèle le moindre développement du *Plasmodium* dans les hématies G6PD déficientes.

La carte des régions du monde où la fréquence des allèles G6PD déficients est supérieure à 1 % recouvre globalement celle où le paludisme sévit de façon endémique. Les principales discordances s'expliquent aisément : le paludisme a été éradiqué depuis une centaine d'années seulement dans le bassin méditerranéen ; il n'y a pas de paludisme en Amérique du Nord mais les personnes G6PD déficientes sont d'origine africaine. Cette corrélation laisse à penser que l'extension des allèles G6PD déficients, dans les régions où le paludisme

est endémique, est due à l'avantage sélectif qui résulte de la possession d'allèle G6PD déficient vis-à-vis du paludisme. Cela est confirmé à l'échelle locale par les données relatives à l'archipel du Vanuatu : alors que le peuplement des îles a la même origine, la fréquence de la déficience en G6PD est beaucoup plus grande dans celle où le paludisme est répandu. Le maintien d'une fréquence relativement forte d'allèles à l'origine d'une G6PD déficiente dans des zones d'endémie palustre résulte d'un équilibre entre l'avantage contre le *Plasmodium* assuré par ces allèles et le désavantage dû à la déficience enzymatique.

Les données épidémiologiques confirment l'avantage sélectif des porteurs d'allèles G6PD déficients dans les régions impaludées.

Filles hétérozygotes G6PDB / G6PDA- ou G6PDA / G6PDA-			
	Non paludéennes	Paludisme modéré	Paludisme sévère
Gambie	13,7%	9,1%	6,7%
Kenya	27,3%	17,2%	18,8%

Fréquence des filles hétérozygotes chez les filles non paludéennes et chez des filles souffrant d'un paludisme plus ou moins sévère

Garçons de génotype G6PDA-			
	Non paludéens	Paludisme modéré	Paludisme sévère
Gambie	5,9%	2,7%	1,4%
Kenya	18,8%	14,3%	11,1%

Fréquence des garçons G6PDA- chez des garçons non paludéens et chez des garçons souffrant d'un paludisme plus ou moins sévère

D'après RUWENDE C. et al. (1995).

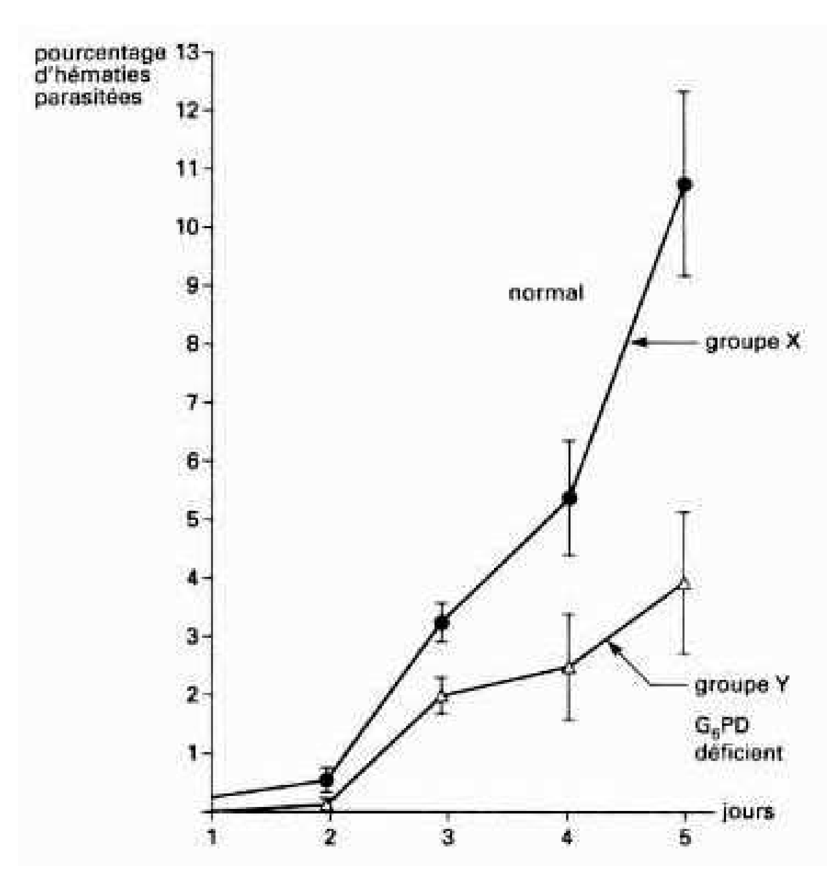
En effet, dans une population **d'enfants paludéens** d'Afrique de l'Ouest et de l'Est, la fréquence des garçons et filles porteurs de l'allèle G6PDA-, est **significativement inférieure** à celle **d'enfants non paludéens**, en particulier chez les enfants victimes d'un paludisme sévère. Tout se passe comme si la possession conférait une résistance vis-à-vis du *Plasmodium*.

À l'échelle cellulaire, cela est confirmé par les données relatives aux femmes hétérozygotes parasitées par le *Plasmodium* : chez ces femmes, la fréquence des hématies G6PD déficientes parasitées est, suivant les individus, 2 à 80 fois plus faible que celle des hématies parasitées possédant l'enzyme G6PD efficace (par suite du phénomène de lyonisation, les allèles des gènes portés par un chromosome X s'expriment dans certaines cellules, les allèles de l'autre chromosome X dans d'autres cellules, d'où la présence de deux populations d'hématies chez les femmes hétérozygotes pour le gène G6PD).

	Individus (Filles)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Hématies G6PD déficiente	1,2	0,2	0,14	5,3	3,6	0,07	0,17	1,2	0,64	2,9	2,4	7,8	5	0,31	0,39	2,3	1,4	1,9	0,03	0,11
Hématies G6PD fonctionnelle	2,3	3,4	1,5	29	12	5,5	2,2	7,3	6,6	6,7	6,4	23	13,7	10,2	2,4	16,5	5,1	14,5	0,28	0,88

D'après LUSATTO L. et al. (1969).

Les données expérimentales fournissent un début d'explication à la résistance vis-à-vis du *Plasmodium* que confère la possession d'un allèle G6PD déficient. D'une part, on constate qu'*in vitro* les hématies G6PD déficientes sont moins parasitées que les hématies non déficientes ; d'autre part, lorsqu'elles sont parasitées, elles sont plus phagocytées à un stade où le parasite n'a pas encore réalisé sa multiplication. Il en résulte une moins grande production de *Plasmodium* et donc un paludisme moins sévère.



Étude expérimentale : comparaison in vitro *du pourcentage d'hématies parasitées chez une population déficiente en G6PD et une population non déficiente*

PROCRÉATION – ANALYSE DE CAS CLINIQUES

Déterminisme du sexe

Informations scientifiques

Chez l'embryon de six semaines, les gonades sont encore indifférenciées et des voies génitales mâles (canaux de Wolf) et féminines (canaux de Müller) sont présentes. C'est le caryotype de l'embryon qui détermine la différenciation des gonades, prélude à la différenciation des voies génitales.

Si l'embryon possède un chromosome Y, le gène SRY (S*ex Determining Region of Y Chromosom*) s'exprime et il y a alors production de protéine TDF (*Testicule Determining Factor*). Cette protéine contrôle l'expression de nombreux autres gènes codant pour des protéines induisant la différenciation des gonades en testicules dès la 8e semaine. Les testicules produisent alors de la testostérone, responsable du développement des voies génitales mâles et de la masculinisation des organes génitaux externes ; ils produisent également de l'AMH (*Anti Mullerian Hormon*) responsable de la dégénérescence des canaux de Müller.

Si l'embryon ne possède pas de chromosome Y, les gonades se différencient en ovaires (vers la 10e semaine). L'absence de testostérone entraîne la régression des canaux de Wolf et l'absence d'AMH entraîne le développement des canaux de Müller, à l'origine des voies génitales féminines.

Le caryotype ne détermine donc que la différenciation des gonades. La différenciation des voies génitales se fait ensuite de façon indépendante du caryotype, en fonction des hormones sécrétées ou non (testostérone et AMH).

L'achèvement des phénotypes sexuels se déroule à la puberté. Les gonades deviennent alors fonctionnelles et produisent des gamètes, mais aussi des hormones gonadiques agissant sur les caractères sexuels secondaires. Les gonatrophines hypophysaires (LH et FSH) contrôlent le fonctionnement de ces gonades ; leur sécrétion est elle-même sous l'influence de la GnRH hypothalamique et sous l'influence du rétrocontrôle exercé par les hormones gonadiques.

Toute altération dans la structure d'une des hormones impliquées (AMH, TDF, gonadotrophines, hormones gonadiques) ou de leurs récepteurs peut donc perturber la mise en place du phénotype sexuel.

Pistes d'exploitation pédagogique des données fournies

Les données fournies peuvent être utilisées en cours d'étude du thème sur le déterminisme du sexe, comme support pour établir certaines notions du programme ; elles peuvent aussi servir de support à un travail de réflexion et de recherche en fin d'étude de ce thème, les élèves ayant alors à mobiliser les connaissances nouvellement acquises. Neuf cas cliniques ont été retenus.

Cas	Phénotype et caryotype	Cause physiologique
Cas n° 1 – Hypogonadisme hypophysaire	Masculin – absence de puberté – 46, XY	LH déficiente
Cas n° 2 – Pseudo-hermaphrodisme	Féminin - absence de puberté – 46, XY	Récepteur à LH déficient
Cas n° 3 – Hypogonadisme gonadique	Féminin – absence de puberté – 46, XX	Récepteur à LH déficient
Cas n° 4 – Hypogonadisme hypophysaire	Masculin – puberté normale – stérile – 46, XY	FSH déficiente
Cas n° 5 – Hypogonadisme hypophysaire	Féminin – absence de puberté – 46, XX	FSH déficiente
Cas n° 6 – Hypogonadisme hypogonadotrope	Masculin – absence de puberté et cryptorchidie – 46, XY	Récepteur à GnRH déficient
Cas n° 7 – Syndrome de Swyer	Féminin – absence de puberté – 46, XY	TDF (protéine SRY) déficient
Cas n° 8 – PMDS (Homme à utérus)	Masculin – présence d'un utérus – 46, XY	AMH déficiente
Cas n° 9 – Testicule féminisant	Féminin – absence de puberté – 46, XY	Récepteur à la testostérone déficient

Pour chaque cas, sont disponibles :

- des données présentées en un seul document : phénotype clinique, caryotype, données histologiques, dosages hormonaux, résultats de traitements et tests ;
- des données moléculaires : séquences des hormones FSHβ, LHβ, AMH, SRY, récepteur à LHβ, récepteur à FSHβ, récepteur à AMH, récepteur à la GnRH, récepteur à la testostérone ;
- les données fournies sur les dosages hormonaux dans les différents cas cliniques sont à comparer avec les valeurs considérées comme physiologiquement normales.

	Femme	Homme
LH en mUI par ml 1 mUI de LH par ml correspond à 1 ng par ml environ.	Folliculaire : 1.5 à 10 Pic ovulatoire : 18 à 90 Lutéale : 1 à 16 Ménopause : 15 à 90	2 à 10
FSH en mUI par ml 1 mUI de FSH par ml correspond à 1 ng par ml environ.	Folliculaire : 2 à 17 Pic ovulatoire : 9 à 26 Lutéale : 2 à 8 Ménopause : 20 à 150	1 à 12
Testostérone en nmol par l en ng par dl (100 ml)	0.8 à 2.6 23 à 75	10 à 38 300 à 1 200
Œstrogène (œstradiol) en pg par ml en pmol par l	Folliculaire : 30 à 90 Pic préovulatoire : 90 à 400 Lutéale : 50 à 20 Ménopause : inférieur à 20 Au cours du cycle : de 90 à 1 470	
Progestérone en ng par ml	Folliculaire : 0.1 à 0.5 Lutéale : 3 à 20 Ménopause : inférieur à 2	
AMH en pmol/l	- Indécelable jusqu'à la puberté - Très faible chez l'adulte	- Taux croissant pendant la période fœtale - Stable jusqu'à la puberté : de 300 à 400 - Très faible à l'âge adulte

Cas n°1 (hypogonadisme hypophysaire)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Procréation - Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : séquences de référence** puis **Hormones déterminisme du sexe** qui charge le fichier sequencesref-cascliniq.edi affichant les séquences strictement codantes de référence (codant pour des protéines fonctionnelles) des allèles « normaux » des gènes AMH, FSHβ, LHβ, RAMH (récepteur à l'AMH), SRY, RFSH (récepteur à la FSH), RLH (récepteur à la LH), RGNRH(récepteur à la GnRH), Rtesto (récepteur à la testostérone) ;
- **Cas 1 : Homme non pubère** puis **Hormones déterminisme du sexe** qui charge le fichier sequences-cas1.edi affichant les séquences strictement codantes des allèles des gènes précédents pour le cas 1.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Procréation - Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : dosages de référence** puis **Normes pour les dosages** qui charge le fichier normes.bmp affichant les normes humaines des taux de quelques hormones impliquées dans la fonction de reproduction (LH, FSH, AMH, œstrogènes, progestérone, testostérone) ;
- **Cas 1 : Homme non pubère** puis **Dosages cas 1** qui charge le fichier cas1.bmp présentant des données concernant le cas n° 1 – données phénotypiques, caryotype, dosages hormonaux, réponses aux tests de stimulation.

Données phénotypiques

Homme de phénotype masculin normal, mais ayant un pénis infantile et des testicules de petite taille ; cet Homme a consulté pour la première fois à 17 ans à cause d'un retard pubertaire.

Traitements : débuté à 17,5 ans – prise de testostérone à raison de 300 mg toutes les deux semaines. Ce traitement a eu pour effet de développer les poils pubiens, la barbe, de faire muer la voix, de faire croître le pénis et se développer les testicules, de permettre les éjaculations.

Arrêt du traitement à 19,5 ans. L'arrêt du traitement a supprimé la capacité d'éjaculer et a fait ralentir la croissance de la barbe.

Caryotype **46, XY**

Données histologiques

La biopsie testiculaire a révélé que les cellules de Leydig dans le tissu interstitiel sont peu apparentes et peu développées.

Dosages hormonaux

(Dosages immunologiques, les dosages par fixation au récepteur spécifique donnent une valeur nulle).

Dosages réalisés de 20 à 25 ans, après arrêt du traitement hormonal :

LH : 18 à 34 UI/l FSH : 8 à 10 UI/l Testostérone : 30 à 80 ng/100 ml

Test de stimulation à la LH

(injection unique d'une certaine dose de LH)

N° du test	Dose de LH injectée en UI	Taux de testostérone en ng/ml
Test 1 Le 24 mars 1973	250	T 0 h = 70
		T 6 h = 210
		T 1 j = 430
		T 2 j = 310
Test 2 Le 7 avril 1973	500	T 0 h = 50
		T 1 j = 510
		T 2 j = 320
Test 3 Le 21 avril 1973	2 000	T 0 h = 30
		T 1 j = 390
		T 2 j = 460
Test 4 Le 19 mai 1973	4 000	T 0 h = 50
		T 1 j = 380
		T 2 j = 650

Test de stimulation par la GnRH

Les mesures suivantes donnent le taux de LH au cours des deux heures qui suivent une injection unique de 100 microgrammes de GnRH par voie intraveineuse.

t=0 : 18 UI/l

t=15 min : 40 UI/l

t=30 min: 60 UI/l

t=2 h : 20 UI/l

Exploitation des données

Le phénotype est conforme au génotype ; il y a donc eu sécrétion et action normale des hormones impliquées dans la différenciation sexuelle au cours du développement fœtal (testostérone et AMH).

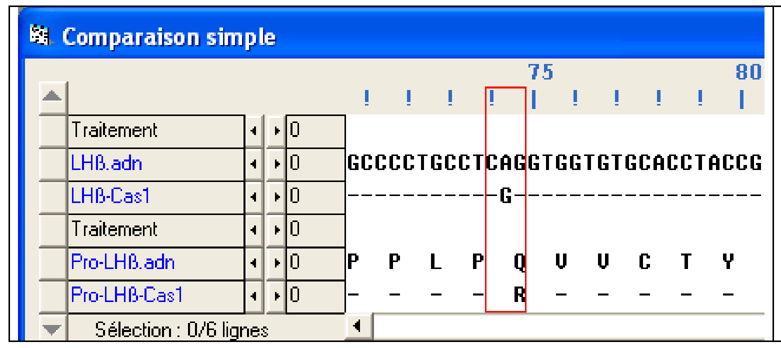
Les organes cibles répondent bien à la testostérone puisqu'un traitement par injection de cette hormone a permis le développement pubertaire.

Les dosages hormonaux montrent un déficit de sécrétion de testostérone, et indiquent une concentration de LH supérieure à la normale. La GnRH n'est donc pas en cause, ni les récepteurs hypophysaires à cette GnRH.

Le fait que le dosage de la LH par fixation à ses récepteurs indique des valeurs quasiment nulles alors que la LH est présente en concentration normale permet d'envisager un problème au niveau de la LH.

Le test par la LH exogène permet de vérifier que le récepteur à la LH est fonctionnel.

La comparaison avec *Anagène* des séquences des différentes hormones et récepteurs de l'individu du cas n° 1 avec les séquences de référence ne montre de différences que pour la LH :



Mutation par substitution du nucléotide A en nucléotide G, modifie le 74e codon (CAG → CGG).

Ainsi, la glutamine 74 est remplacée par l'arginine.

L'acide aminé 74 étant impliqué dans la fixation de l'hormone à son récepteur (sur les cellules de Leydig), la mutation empêche cette fixation.

Remarques

Cet exemple montre que la LH n'est pas nécessaire à la différenciation de l'appareil génital masculin au cours de la vie fœtale. Cependant, les cellules de Leydig fœtales doivent être stimulées pour produire la testostérone indispensable à cette différenciation. C'est l'HCG placentaire qui passe dans la circulation fœtale qui joue ce rôle.

La thérapeutique envisageable est à base d'injections régulières de testostérone ou d'HCG (hormone ayant les mêmes cibles que la LH, mais présentant un effet plus prolongé grâce à une demi-vie plus longue).

Cas n°2 (hypogonadisme chez deux filles)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : séquences de référence** puis **Hormones déterminisme du sexe** qui charge le fichier sequencesref-cascliniq.edi affichant les séquences strictement codantes de référence (codant pour des protéines fonctionnelles) des allèles « normaux » des gènes AMH, FSHβ, LHβ, RAMH (récepteur à l'AMH), SRY, RFSH (récepteur à la FSH), RLH (récepteur à la LH), RGNRH (récepteur à la GnRH), Rtesto (récepteur à la testostérone) ;
- **Cas 2 : Femme non pubère** puis **Hormones déterminisme du sexe** qui charge le fichier sequences-cas2.edi affichant les séquences strictement codantes des allèles des gènes précédents pour le cas 2.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : dosages de référence** puis **Normes pour les dosages** qui charge le fichier normes.bmp affichant les normes humaines des taux de quelques hormones impliquées dans la fonction de reproduction (LH, FSH, AMH, œstrogènes, progestérone, testostérone) ;
- **Cas 2 : Femme non pubère** puis **Dosages cas 2** qui charge le fichier cas2.bmp présentant des données concernant le cas n° 2 – données phénotypiques, caryotype, dosages hormonaux, réponses aux tests de stimulation.

Données phénotypiques

Deux sœurs de 37 et 42 ans non pubères : organes génitaux externes féminins, pas de développement des seins, absence de menstruation, absence d'organes génitaux internes féminins (ovaires, trompes, utérus) à l'exception d'un vagin court et fermé ; présence de deux testicules en position interne (abdominales) accompagnés de deux épididymes et de deux canaux déférents atrophiés.

Caryotype

46, XY

Données histologiques

Les testicules ont été enlevés (risque de cancer) et présentent des tubes séminifères sans signes de spermatogenèse. Leur tissu interstitiel ne contient que des cellules de Leydig immatures.

Dosages hormonaux

Testostérone (nmol/l)	Individu 1 : 0,69 Individu 2 : 0,53
LH (UI/l)	Individu 1 : 23,3 Individu 2 : 19

Test de stimulation de la fonction gonadique par l'HCG (équivalent LH)

L'injection intramusculaire de 6 000 UI de HCG ne provoque aucune élévation du taux de testostérone plasmatique.

Test de stimulation de la fonction hypophysaire par la GnRH

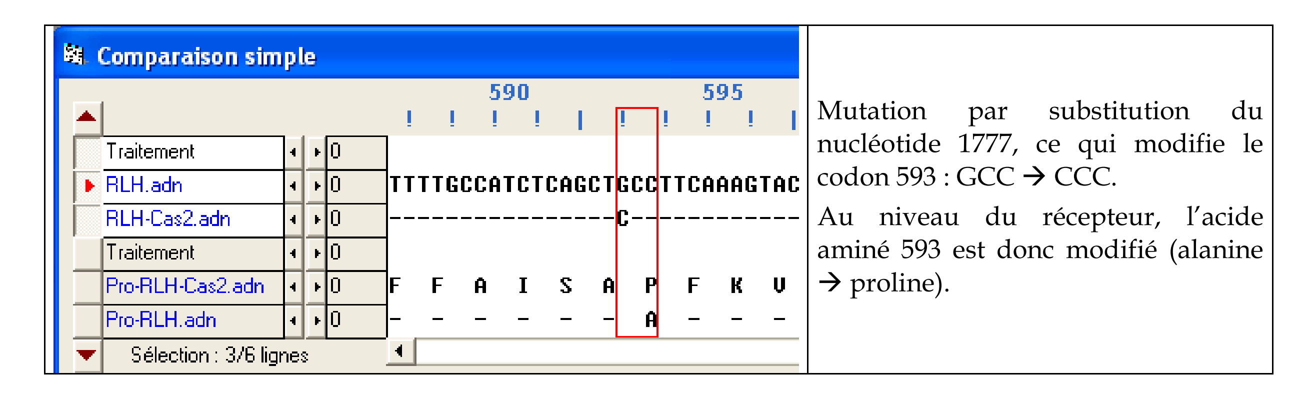
L'injection de 100 microgrammes de GnRH provoque une augmentation importante de la sécrétion de LH et l'élévation normale du taux de FSH.

Exploitation des données

Le phénotype sexuel externe n'est pas conforme au génotype ; il y a donc eu un problème lors de la différenciation sexuelle au cours du développement fœtal. La présence de testicules indique que le gène SRY s'est exprimé normalement. L'absence de voies génitales fonctionnelles indique que l'AMH a été secrétée normalement.

Le test de stimulation gonadique par la HCG (équivalent LH) montre qu'il n'y a pas de réponse hormonale des gonades.

La comparaison avec *Anagène* des séquences des différentes hormones et récepteurs de l'individu du cas n° 2 avec les séquences de référence, ne montre de différences que pour le récepteur à la LH :



Mutation par substitution du nucléotide 1777, ce qui modifie le codon 593 : GCC → CCC.

Au niveau du récepteur, l'acide aminé 593 est donc modifié (alanine → proline).

Remarques

En fait, le récepteur à la LH est capable de fixer l'hormone, mais cette fixation est incapable de déclencher une réponse biologique dans la cellule cible (pas d'augmentation du taux d'AMPc).

Au cours du développement sous l'action du gène SRY, les gonades se sont différenciées en testicules ; ces testicules ont sécrété de l'AMH qui a provoqué la régression des canaux de Müller. Cependant, à cause de la déficience des récepteurs à la LH, l'hormone HCG n'a pas pu entraîner la production de testostérone et les organes génitaux externes n'ont pas pu être masculinisés.

Les données histologiques montrent l'absence de cellules de Leydig, ce qui contribue à la faiblesse du taux de testostérone et met en évidence le rôle important joué par la LH dans la différenciation de ces cellules.

Cas n°3 (hypogonadisme chez une jeune fille)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : séquences de référence** puis **Hormones déterminisme du sexe** qui charge le fichier sequencesref-cascliniq.edi affichant les séquences strictement codantes de référence (codant pour des protéines fonctionnelles) des allèles « normaux » des gènes AMH, FSHβ, LHβ, RAMH (récepteur à l'AMH), SRY, RFSH (récepteur à la FSH), RLH (récepteur à la LH), RGNRH (récepteur à la GnRH), Rtesto (récepteur à la testostérone) ;
- **Cas 3 : Femme non pubère** puis **Hormones déterminisme du sexe** qui charge le fichier sequences-cas3.edi affichant les séquences strictement codantes des allèles des gènes précédents pour le cas 3.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : dosages de référence** puis **Normes pour les dosages** qui charge le fichier normes.bmp affichant les normes humaines des taux de quelques hormones impliquées dans la fonction de reproduction (LH, FSH, AMH, œstrogènes, progestérone, testostérone) ;
- **Cas 3 : Femme non pubère** puis **Dosages cas 3** qui charge le fichier cas3.bmp présentant des données concernant le cas n° 3 – données phénotypiques, caryotype, dosages hormonaux, réponses aux tests de stimulation.

Données phénotypiques

Jeune femme de 18 ans, de phénotype féminin normal. Présence de signes pubertaires : développement normal des seins et des poils pubiens et axillaires, mais absence de menstruations. L'échographie révèle deux ovaires de taille normale.

Caryotype

46, XX

Données histologiques

Absence de corps jaunes et de follicules mûrs dans les ovaires.

Dosages hormonaux

(Gamme des valeurs obtenues par des dosages journaliers pendant un mois)

Les taux mesurés ne présentent aucune variation cyclique quelle que soit l'hormone étudiée.

Aucun changement des concentrations des hormones de FSH ou de LH.

LH	14 – 38 UI/l (moyenne 26)
FSH	8 – 11 UI/l (moyenne 9)
Œstrogènes	6 – 53 pg/ml (moyenne 37)
Progestérone	0,26 – 0,48 ng/ml (moyenne 0,37)

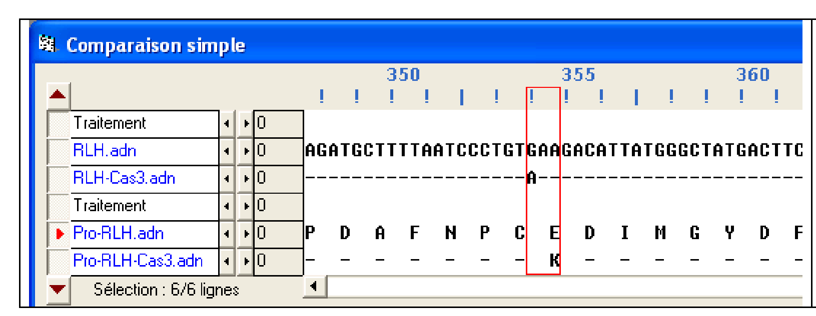
Exploitation des données

Les taux d'hormones hypophysaires sont à peu près normaux, mais sans variation cyclique ; par contre, les taux d'hormones ovariennes sont très faibles.

Les follicules se développent, mais l'ovulation n'a pas lieu, et aucun corps jaune ne se forme donc, expliquant le taux très faible de progestérone. La FSH doit donc agir normalement, par contre il y a défaillance au niveau de l'action de la LH

Deux hypothèses sont alors possibles : structure anormale de la LH ou structure anormale de son récepteur.

La comparaison avec *Anagène* des séquences des différentes hormones et récepteurs de l'individu du cas n° 3 avec les séquences de référence, ne montre de différences que pour le récepteur à la LH :



Mutation par substitution du nucléotide 1060, ce qui modifie le codon 354 : GAA → AAA.

Au niveau du récepteur, l'acide aminé 354 est donc modifié (acide glutamique → lysine).

Remarques

Le taux élevé de LH s'explique par l'absence de rétrocontrôle de la part des œstrogènes (taux trop faible) et de la progestérone (taux quasiment nul) sur l'axe hypothalamo-hypophysaire

Cas n°4 (hypogonadisme chez un jeune Homme)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : séquences de référence** puis **Hormones déterminisme du sexe** qui charge le fichier sequencesref-cascliniq.edi affichant les séquences strictement codantes de référence (codant pour des protéines fonctionnelles) des allèles « normaux » des gènes AMH, FSHβ, LHβ, RAMH (récepteur à l'AMH), SRY, RFSH (récepteur à la FSH), RLH (récepteur à la LH), RGNRH (récepteur à la GnRH), Rtesto (récepteur à la testostérone) ;
- **Cas 4 : Homme stérile** puis **Hormones déterminisme du sexe** qui charge le fichier sequences-cas4.edi affichant les séquences strictement codantes des allèles des gènes précédents pour le cas 4.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : dosages de référence** puis **Normes pour les dosages** qui charge le fichier normes.bmp affichant les normes humaines des taux de quelques hormones impliquées dans la fonction de reproduction (LH, FSH, AMH,œstrogènes, progestérone, testostérone) ;
- **Cas 4 : Homme stérile** puis **Dosages cas 4** qui charge le fichier cas4.bmp présentant des données concernant le cas n° 4 – données phénotypiques, caryotype, dosages hormonaux, réponses aux tests de stimulation.

Données phénotypiques

Homme de 28 ans, de phénotype masculin pubère, qui présente une libido et des capacités sexuelles normales.

Les organes génitaux externes sont normaux, mais les deux testicules sont petits et mous. Cet Homme n'a pas eu d'enfants de deux mariages successifs.

Caryotype

46, XY

Données histologiques

L'analyse du sperme met en évidence une azoospermie.

Une biopsie a montré la présence de cellules de Leydig normales.

Dosages hormonaux

LH	12 UI/l
FSH	Indétectable (par deux méthodes différentes)
Testostérone	11 nmol/l

Test de stimulation de la fonction hypophysaire par injection de 100 microgrammes de GnRH

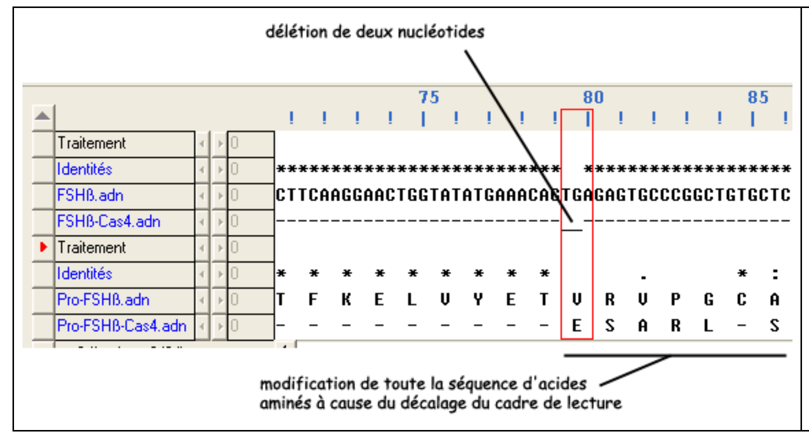
LH	41 UI/l
FSH	Indétectable (par deux méthodes différentes)

Exploitation des données

On constate un taux à peu près normal de LH et de testostérone, mais une absence de sécrétion de FSH. Ce manque de FSH explique le non-déroulement de la spermatogenèse.

L'absence de FSH malgré une stimulation par la GnRH permet de penser que le problème se situe au niveau de la FSH elle-même (les récepteurs de la GnRH au niveau des cellules hypophysaires ne sont pas en cause car la stimulation des cellules sécrétrices de LH se fait normalement).

La comparaison avec *Anagène* (alignement avec discontinuités) des séquences des différentes hormones et récepteurs de l'individu du cas n° 4 avec les séquences de référence, ne montre de différences que pour la FSH :



La délétion de deux nucléotides dans la séquence codante du gène de la FSH entraîne un changement important de la séquence d'acides aminés à partir du n° 79.

Cela entraîne l'apparition précoce d'un codon stop (codon n° 105) et donc la synthèse d'une protéine plus courte de 104 acides aminés au lieu de 129.

Remarques

Le taux de testostérone se situe dans la norme basse, ce qui laisse penser que la FSH aurait aussi un rôle sur les cellules de Leydig.

Le taux élevé de LH s'explique par la faiblesse du rétrocontrôle négatif de la testostérone assez basse.

Cas n°5 (hypogonadisme chez une jeune fille)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : séquences de référence** puis **Hormones déterminisme du sexe** qui charge le fichier sequencesref-cascliniq.edi affichant les séquences strictement codantes de référence (codant pour des protéines fonctionnelles) des allèles « normaux » des gènes AMH, FSHβ, LHβ, RAMH (récepteur à l'AMH), SRY, RFSH (récepteur à la FSH), RLH (récepteur à la LH), RGNRH (récepteur à la GnRH), Rtesto (récepteur à la testostérone) ;
- **Cas 5 : Femme non pubère** puis **Hormones déterminisme du sexe** qui charge le fichier sequences-cas5.edi affichant les séquences strictement codantes des allèles des gènes précédents pour le cas 5.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : dosages de référence** puis **Normes pour les dosages** qui charge le fichier normes.bmp affichant les normes humaines des taux de quelques hormones impliquées dans la fonction de reproduction (LH, FSH, AMH, Oestrogènes, progestérone, testostérone) ;
- **Cas 5 : Femme non pubère** puis **Dosages cas 5** qui charge le fichier cas5.bmp présentant des données concernant le cas n° 5 – données phénotypiques, caryotype, dosages hormonaux, réponses aux tests de stimulation.

Données phénotypiques

Jeune fille de 16 ans, de phénotype féminin normal, qui présente des signes pubertaires comme le développement des poils pubiens, mais pas de développement des seins et une absence de menstruations. Cette jeune fille présente un âge osseux de 14 ans. L'échographie révèle deux ovaires de taille normale.

Le développement des seins a été induit par un traitement œstrogénique.

Caryotype

46, XX

Données histologiques

Une biopsie des ovaires montre l'absence de follicules cavitaires développés.

Dosages hormonaux

(Sur une période d'un mois, les concentrations mesurées ne présentent aucune variation cyclique, quelle que soit l'hormone étudiée.)

LH	21 UI/l
FSH	< 0,5 UI/l
Œstrogènes (œstradiol)	25 pg/ml

Test de stimulation de la fonction hypophysaire par injection de 100 microgrammes de GnRH

Hormones dosées	Taux initial	Taux au bout de 30 minutes	Taux au bout de 60 minutes
LH (UI/l)	33	170	130
FSH (UI/l)	0,6	0,6	0,8

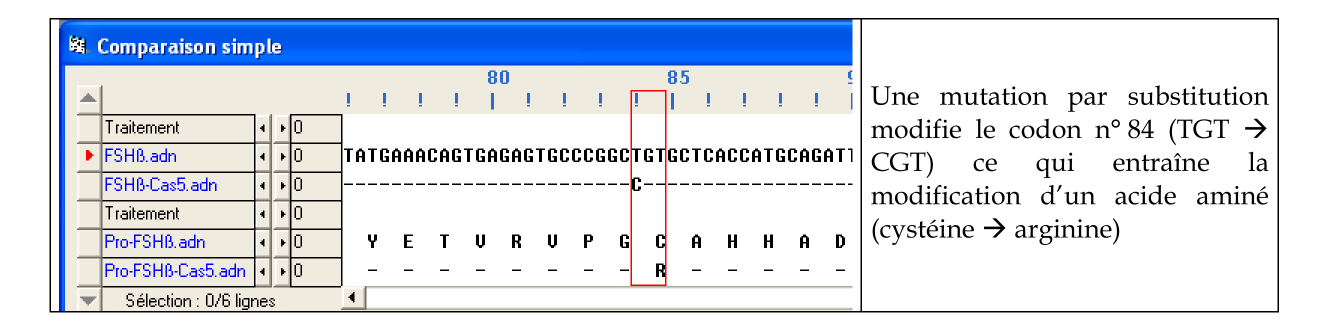
Exploitation des données

On constate un taux très faible de FSH, et un taux d'œstrogènes correspondant pratiquement à celui de la ménopause ; cela est à mettre en relation avec un cycle ovarien incomplet.

La stimulation par la GnRH reste sans effet sur le taux de FSH.

Il semble donc que la LH soit active, et sécrétée en réponse à la stimulation par la GnRH, contrairement à la FSH. On peut donc envisager un problème au niveau de la FSH.

La comparaison avec *Anagène* des séquences des différentes hormones et récepteurs de l'individu du cas n° 5 avec les séquences de référence, ne montre de différences que pour la FSH :



Remarques

Le taux élevé de LH s'explique par la faiblesse du rétrocontrôle négatif des hormones ovariennes (absence de progestérone et taux très faible d'œstrogènes).

La comparaison des cas 3 et 5 montre une similitude des effets d'une inactivité de la LH et de la FSH en ce qui concerne la production d'œstrogènes. Les deux gonadotrophines sont donc indispensables à cette production, leurs actions se complètent (la FSH permet la croissance du follicule dont les cellules doivent être stimulées par la LH pour produire des œstrogènes. Sous l'action de la LH, il y a production d'androgènes par les cellules de la thèque qui sont convertis en œstrogènes par les cellules de la granulosa sous l'action de la FSH).

Cas n°6 (hypogonadisme chez un jeune Homme)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : séquences de référence** puis **Hormones déterminisme du sexe** qui charge le fichier sequencesref-cascliniq.edi affichant les séquences strictement codantes de référence (codant pour des protéines fonctionnelles) des allèles « normaux » des gènes AMH, FSHβ, LHβ, RAMH (récepteur à l'AMH), SRY, RFSH (récepteur à la FSH), RLH (récepteur à la LH), RGNRH (récepteur à la GnRH), Rtesto (récepteur à la testostérone) ;
- **Cas 6 : Homme non pubère** puis **Hormones déterminisme du sexe** qui charge le fichier sequences-cas6.edi affichant les séquences strictement codantes des allèles des gènes précédents pour le cas 6.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : dosages de référence** puis **Normes pour les dosages** qui charge le fichier normes.bmp affichant les normes humaines des taux de quelques hormones impliquées dans la fonction de reproduction (LH, FSH, AMH, œstrogènes, progestérone, testostérone) ;
- **Cas 6 : Homme non pubère** puis **Dosages cas 6** qui charge le fichier cas6.bmp présentant des données concernant le cas n° 6 – données phénotypiques, caryotype, dosages hormonaux, réponses aux tests de stimulation.

Données phénotypiques

Garçon de 19 ans sans pilosité faciale et thoracique et sans autres signes pubertaires. Ses deux testicules sont en position abdominale et de petite taille.

Un scanner de la région hypothalamo-hypophysaire n'a permis de déceler aucune anomalie.

Caryotype

46, XY

Données histologiques

Les testicules ne présentent pas de spermatogenèse complète (absence notamment de spermatozoïdes).

Dosages hormonaux

LH	< 0,9 UI/l
FSH	< 0,4 UI/l
Testostérone	< 0,7 nmol/ml

Test de stimulation de la fonction hypophysaire par la GnRH

Administration de GnRH de façon pulsatile ; une pulse toutes les deux heures, pendant 2 mois.

	Dose de GnRH 60 ng/kg toutes les 2 heures	Dose de GnRH 60 ng/kg toutes les 2 heures
LH (UI/l)	< 0,5 UI/l	< 0,5 UI/l
FSH (UI/l)	< 0,2 UI/l	< 0,2 UI/l
Testostérone	3,7 nmom/l	2,7 nmom/l

Traitement par une association de hMG (75 UI) – HCG (500 UI) trois fois par semaine

Après quatorze mois de traitement, la biopsie testiculaire montre une spermatogenèse complète (formation de spermatozoïdes).

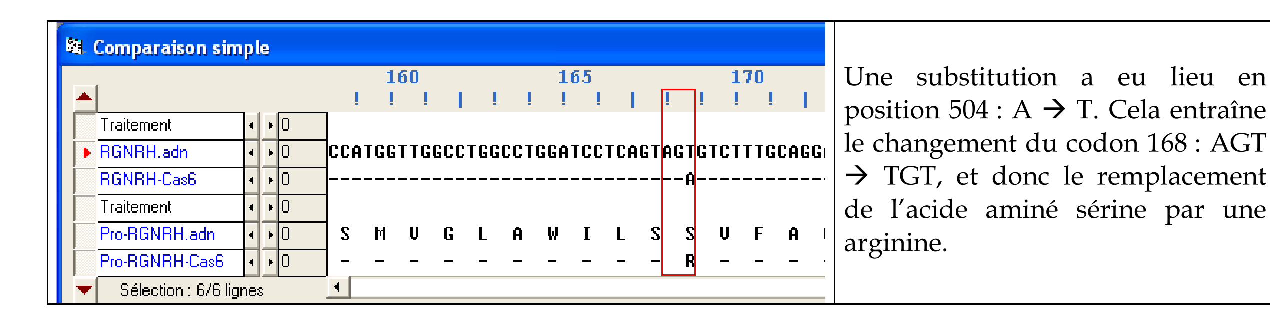
Durée du traitement	Deux mois	Dix mois	Vingt-quatre mois
Testostérone	16,8 nmol/l	27,1 nmol/l	24 nmol/l
Volume des testicules (en ml)	Gauche : 10 Droit : 10	Gauche : 12 Droit : 10	Gauche : 15 Droit : 14

Exploitation des données

On observe un taux très faible des hormones FSH, LH et testostérone, et ce même en cas de stimulation par la GnRH. Ces taux très faibles expliquent l'absence de spermatogenèse complète.

Le traitement par la HMG-HCG (équivalent LH) est par contre efficace ; les récepteurs à ces hormones sont donc fonctionnels. On peut alors envisager un problème au niveau de la GnRH ou de ses récepteurs.

La comparaison avec *Anagène* des séquences des différentes hormones et récepteurs de l'individu du cas n° 6 avec les séquences de référence ne montre de différences que pour les récepteurs à la GnRH :



Une substitution a eu lieu en position 504 : A → T. Cela entraîne le changement du codon 168 : AGT → TGT, et donc le remplacement de l'acide aminé sérine par une arginine.

Remarques

Il n'y a pas eu d'anomalie lors de la différenciation des testicules durant la vie fœtale car l'HCG placentaire produite (de façon indépendante du complexe hypothalamo-hypophysaire) a permis une différenciation des organes génitaux dans le sens mâle.

Cas n°7 (inversion sexuelle chez deux filles)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : séquences de référence** puis **Hormones déterminisme du sexe** qui charge le fichier sequencesref-cascliniq.edi affichant les séquences strictement codantes de référence (codant pour des protéines fonctionnelles) des allèles « normaux » des gènes AMH, FSHβ, LHβ, RAMH (récepteur à l'AMH), SRY, RFSH (récepteur à la FSH), RLH (récepteur à la LH), RGNRH (récepteur à la GnRH), Rtesto (récepteur à la testostérone) ;
- **Cas 7 : Femme non pubère (Syndrome de Swyer)** puis **Hormones déterminisme du sexe** qui charge le fichier sequences-cas7.edi affichant les séquences strictement codantes des allèles des gènes précédents pour le cas 7.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : dosages de référence** puis **Normes pour les dosages** qui charge le fichier normes.bmp affichant les normes humaines des taux de quelques hormones impliquées dans la fonction de reproduction (LH, FSH, AMH, œstrogènes, progestérone, testostérone) ;
- **Cas 7 : Femme non pubère (Syndrome de Swyer)** puis **Dosages cas 7** qui charge le fichier cas7.bmp présentant des données concernant le cas n° 7 – données phénotypiques, caryotype, dosages hormonaux, réponses aux tests de stimulation.

Données phénotypiques

Deux femmes de 20 et 25 ans, ayant des organes génitaux externes féminins normaux. Absence de menstruation, seins peu développés. Les gonades apparaissent à la cœlioscopie comme une bandelette nacrée fibreuse. Les trompes et l'utérus sont de morphologie normale, mais de petite taille.

Caryotype

46, XY

Données histologiques

Les biopsies réalisées ont montré l'absence de différenciation, chez les deux sœurs, des gonades gauche et droite.

Dosages hormonaux

LH	6 UI/l
FSH	35 UI/l
Œstrogènes (œstradiol)	20 pg/ml

Tests : après stimulation hormonale par des injections pulsatiles de GnRH, les concentrations de LH et FSH ont été remesurées

(valeurs mesurées 60 minutes après injection de GnRH).

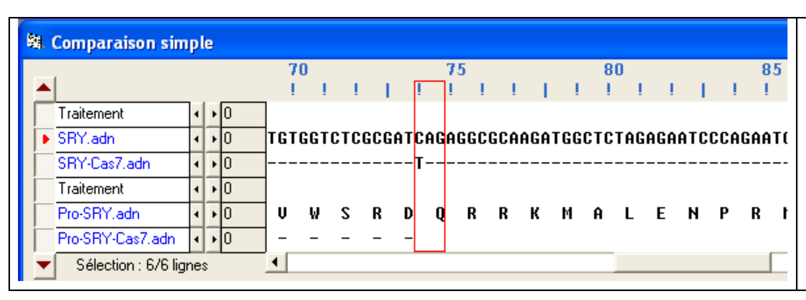
LH	56 UI/l
FSH	68 UI/l

Exploitation des données

Il y a contradiction entre le sexe phénotypique externe (féminin) et le caryotype (masculin).

L'absence de différenciation des gonades permet d'envisager un problème au niveau du gène SRY (codant pour le facteur tdf responsable de la différenciation des gonades en testicules)

La comparaison avec *Anagène* des séquences des différentes hormones et récepteurs de l'individu du cas n° 7 avec les séquences de référence, ne montre de différences que pour le gène SRY :



Une mutation par substitution au niveau du codon 74 (remplacement C → A) entraîne le changement de ce codon (CAG → TAG). On a donc apparition d'un codon stop précoce. La protéine synthétisée est alors beaucoup plus courte : 73 acides aminés au lieu de 205.

Remarques

Il n'y a pas eu de sécrétion d'AMH ni de testostérone pendant le développement fœtal, donc pas de différenciation des gonades en testicules ; il n'y a pas eu non plus différenciation en ovaires, par manque d'un deuxième chromosome X.

Les concentrations en hormones gonadiques sont donc très faibles, et celles en gonadotrophines comparativement élevées, par manque de rétrocontrôle négatif.

Cas n°8 (Homme à utérus)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : séquences de référence** puis **Hormones déterminisme du sexe** qui charge le fichier sequencesref-cascliniq.edi affichant les séquences strictement codantes de référence (codant pour des protéines fonctionnelles) des allèles « normaux » des gènes AMH, FSHβ, LHβ, RAMH (récepteur à l'AMH), SRY, RFSH (récepteur à la FSH), RLH (récepteur à la LH), RGNRH (récepteur à la GnRH), Rtesto (récepteur à la testostérone) ;

- **Cas 8 : Homme à utérus** puis **Hormones déterminisme du sexe** qui charge le fichier sequences-cas8.edi affichant les séquences strictement codantes des allèles des gènes précédents pour le cas 8.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : dosages de référence** puis **Normes pour les dosages** qui charge le fichier normes.bmp affichant les normes humaines des taux de quelques hormones impliquées dans la fonction de reproduction (LH, FSH, AMH, œstrogènes, progestérone, testostérone) ;

- **Cas 8 : Homme à utérus** puis **Dosages cas 8** qui charge le fichier cas8.bmp présentant des données concernant le cas n° 8 – données phénotypiques, caryotype, dosages hormonaux, réponses aux tests de stimulation.

Données phénotypiques

Garçon de 10 ans, de taille normale, présentant des organes génitaux externes normaux mais des testicules non palpables au niveau des bourses.

L'observation de la cavité abdominale a révélé la présence de deux testicules, dans la région pelvienne, entourés chacun d'une trompe de Fallope. Les testicules ont été descendus manuellement ; de taille normale, ils sont connectés à des canaux déférents.

On a également constaté, dans la partie médiane de l'abdomen, la présence d'un utérus de petite taille.

Au cours de son suivi médical, le patient a présenté des signes pubertaires normaux pour son âge.

Caryotype

46, XY

Données histologiques

Une biopsie testiculaire montre un aspect normal des tubes séminifères.

Dosages hormonaux

Le taux de testostérone, bas et fluctuant chez l'enfant, a été mesuré après stimulation par l'HCG (analogue LH) à raison de 1 500 UI deux fois par semaine pendant trois semaines.

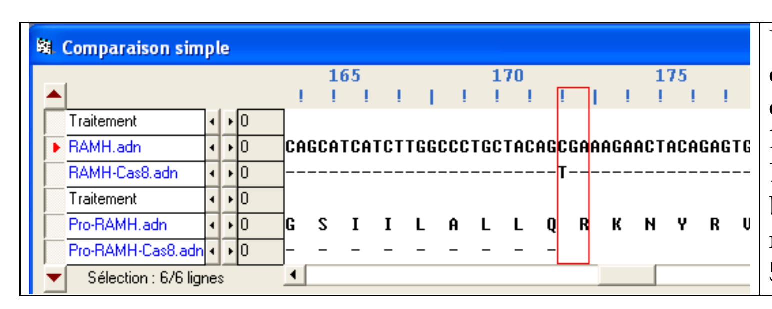
Testostérone après stimulation	30 nmol/l
AMH	350 pmol/l

Exploitation des données

Il y a eu différenciation normale des gonades en testicules, mais pas régression des voies génitales féminines. On peut donc envisager un problème au niveau de l'AMH ou des récepteurs à l'AMH.

La concentration en AMH étant normale pour un enfant de 10 ans laisse à penser qu'elle a été sécrétée durant la vie fœtale mais qu'elle n'a pas agi. Deux explications sont possibles : soit elle est anormale, soit les récepteurs ne sont pas fonctionnels.

La comparaison avec *Anagène* des séquences des différentes hormones et récepteurs de l'individu du cas n° 8 avec les séquences de référence, ne montre de différences que pour le gène des récepteurs à l'AMH :



Une mutation par substitution dans le codon 172 a entraîné la modification de ce codon (CGA → TGA) et donc l'apparition précoce d'un codon stop. La protéine synthétisée est donc beaucoup plus courte que la protéine normale (171 acides aminés au lieu de 573).

Remarques

La déficience des récepteurs à AMH explique que, malgré une production d'AMH normale, ce garçon présente un utérus et une trompe dus à l'absence de régression des canaux de Müller durant le développement fœtal.

Cas n°9 (testicule féminisant)

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : séquences de référence** puis **Hormones déterminisme du sexe** qui charge le fichier sequencesref-cascliniq.edi affichant les séquences strictement codantes de référence (codant pour des protéines fonctionnelles) des allèles « normaux » des gènes AMH, FSHβ, LHβ, RAMH (récepteur à l'AMH), SRY, RFSH (récepteur à la FSH), RLH (récepteur à la LH), RGNRH (récepteur à la GnRH), Rtesto (récepteur à la testostérone) ;
- **Cas 9 : Femme non pubère** puis **Hormones déterminisme du sexe** qui charge le fichier sequences-cas9.edi affichant les séquences strictement codantes des allèles des gènes précédents pour le cas 9.

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Procréation – Analyse de cas cliniques** permet d'atteindre :

- **Cas normal : dosages de référence** puis **Normes pour les dosages** qui charge le fichier normes.bmp affichant les normes humaines des taux de quelques hormones impliquées dans la fonction de reproduction (LH, FSH, AMH, œstrogènes, progestérone, testostérone) ;
- **Cas 9 : Femme non pubère** puis **Dosages cas 9** qui charge le fichier cas9.bmp présentant des données concernant le cas n° 9 – données phénotypiques, caryotype, dosages hormonaux, réponses aux tests de stimulation.

Données phénotypiques

Femme de 29 ans ayant des organes génitaux externes féminins. Le développement des seins est quasi normal, mais elle présente une pilosité axillaire et pubienne peu développée et une absence de menstruations.

Un examen du petit bassin met en évidence la présence d'un vagin court (4 cm) et l'absence d'utérus et de trompes. Deux masses fermes de 3 cm de diamètre (testicules) sont palpables au niveau inguinal.

Chez cette patiente, à l'âge de 14 ans, les seins ont commencé à se développer, mais aucun autre signe pubertaire n'est apparu.

Caryotype

46, XY

Données histologiques

L'analyse des masses détectées dans la région inguinale indique qu'il s'agit de testicules renfermant des cellules de Leydig, des cellules de Sertoli et des spermatogonies, mais aucun stade cellulaire plus avancé de la spermatogenèse (testicule cryptorchide).

Dosages hormonaux

LH	38 UI/l
FSH	52 UI/l
Testostérone	1 350 ng/dl

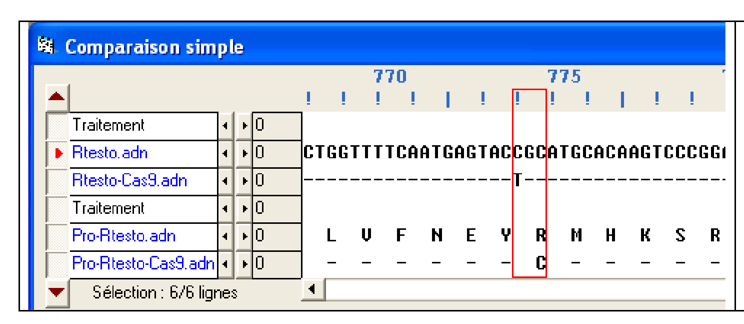
Exploitation des données

Il y a contradiction entre le caryotype et le phénotype sexuel externe, mais pas entre le caryotype et le type de gonades que possède l'individu. Le tractus génital n'a donc pas été masculinisé, bien qu'il y ait eu différenciation des gonades en testicules.

Les taux de gonadotrophines sont très élevés, ainsi que celui de testostérone.

Le tractus génital n'ayant pas été différencié et la spermatogenèse n'étant pas complète, on peut envisager un problème au niveau des récepteurs à la testostérone.

La comparaison avec *Anagène* des séquences des différentes hormones et récepteurs de l'individu du cas n° 9 avec les séquences de référence ne montre de différences que pour le gène des récepteurs à la testostérone :



Une mutation par substitution au niveau du codon 774 a entraîné le changement de ce codon (CGC → TGC), ce qui a entraîné le changement de l'acide aminé correspondant : arginine → cystéine

(cette mutation se situe dans un domaine du récepteur impliqué dans sa liaison à l'androgène).

Remarques

Les testicules ont pu se différencier pendant la vie fœtale sous l'influence du facteur tdf (codé par le gène SRY) ; ils présentent des cellules de Leydig actives produisant la testostérone, mais celle-ci ne peut agir car les récepteurs à la testostérone ne sont pas fonctionnels. Les cellules germinales ne peuvent donc pas être stimulées et assurer la spermatogenèse ; le rétrocontrôle négatif ne peut avoir lieu non plus, ce qui explique le taux élevé de gonadotrophines.

IMMUNOLOGIE

La spécificité des immunoglobulines et des récepteurs T

Informations scientifiques

L'infection par le VIH entraîne des réactions immunitaires de l'organisme qui se traduisent par la production d'anticorps (IgG dirigés contre des peptides de l'enveloppe virale ou de l'enveloppe protéique protégeant le matériel génétique du virus) et la production de LT cytotoxiques (capables de se fixer sélectivement sur des peptides viraux présentés par les cellules infectées, entraînant la lyse de ces cellules).

Les immunoglobulines et les récepteurs T sont des molécules protéiques capables de se lier spécifiquement à un déterminant antigénique. Cependant, si les immunoglobulines sont capables de « reconnaître » directement l'antigène, les récepteurs T ne peuvent le « reconnaître » que s'il est présenté par les molécules HLA de l'organisme (double spécificité).

Les immunoglobulines

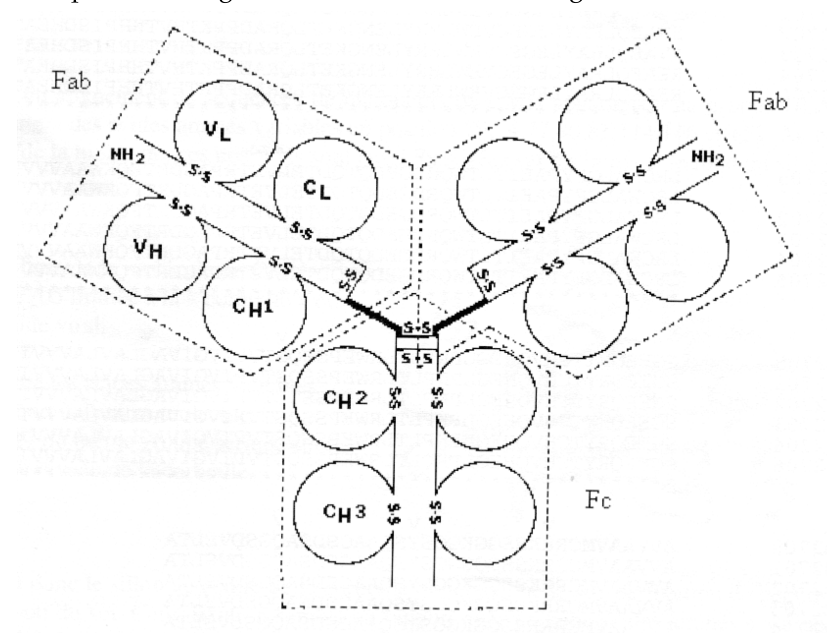
La molécule d'immunoglobuline est très volumineuse (1 336 acides aminés) ; c'est un dimère dont chaque monomère est constitué d'une chaîne lourde (H), de 440 à 450 acides aminés environ, et d'une chaîne légère (L), de 200 à 220 acides aminés environ. Ces chaînes sont organisées en domaines, constitués chacun de 110 à 120 acides aminés. Chaque chaîne lourde comporte quatre domaines globulaires réunis par des petits segments de liaison : trois domaines constants (CH) pour toutes les immunoglobulines et un domaine variable (VH) d'une immunoglobuline à l'autre. Chaque chaîne légère est constituée d'un domaine constant (CL) et d'un domaine variable (VL).

Des ponts disulfures solidarisent l'ensemble de la molécule.

La molécule d'immunoglobuline a été cristallisée en deux morceaux : les deux fragments Fab (*Fragment Antigen Binding*) d'une part, le fragment Fc (*Fragment Cristallisable*) d'autre part. Elle possède une certaine flexibilité entre les fragments Fab et Fc ce qui permet l'écartement entre les deux sites de fixation antigénique en fonction de l'encombrement de l'antigène.

Chaque molécule d'immunoglobuline possède deux sites de fixation antigénique identiques, aux extrémités des deux fragments Fab.

Le document ci-dessous présente l'organisation d'une molécule d'IgG :

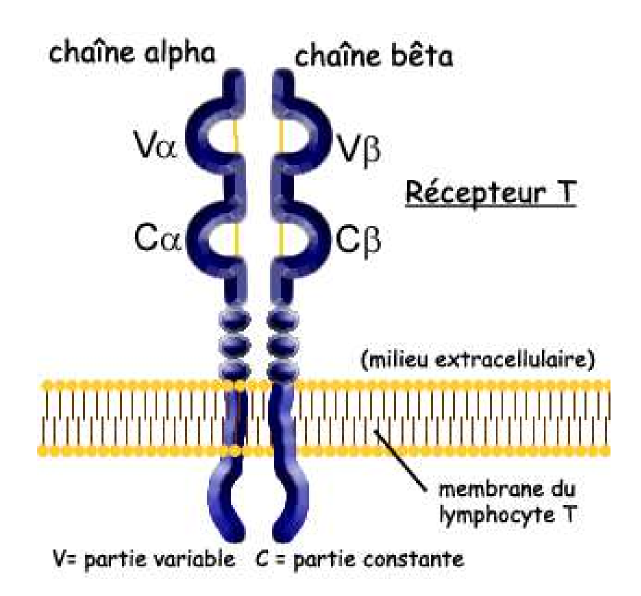


Les immunoglobulines interviennent dans la reconnaissance de l'antigène à deux niveaux :

- cellulaire : les lymphocytes B possèdent des récepteurs membranaires constitués par des IgM monomériques ; les IgM comportent un type de chaîne lourde (μ) qui possède, par rapport à la chaîne lourde γ des IgG, un domaine CH4 supplémentaire prolongé par une queue polypeptidique qui permet un ancrage dans la membrane ;
- plasmatique : les lymphocytes B, transformés en plasmocytes, libèrent des IgG dans le plasma.

Les récepteurs T des lymphocytes T

Un récepteur T est un dimère constitué d'une chaîne alpha et d'une chaîne bêta. Chacune de ces chaînes est constituée d'un domaine constant d'un récepteur T à l'autre, et d'un domaine qui varie selon les récepteurs T et qui est donc responsable de la spécificité.



Pistes d'exploitation pédagogique des données fournies

L'utilisation combinée du logiciel *RasTop* (*fichiers 1F58.pdb, ACY.pdb, 1E6J.pdb, 1NLD.pdb et 1BD2b.pdb*) permet de visualiser les molécules étudiées avec *Anagène* et de mener une démarche pédagogique plus complète.

Bases moléculaires de la spécificité spatiale des immunoglobulines

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Immunologie/IgG/IgG anti VIH** permet d'atteindre **4 IgG anti ViH** qui charge le fichier iggvih.edi affichant les séquences protéiques des quatre chaînes lourdes et légères des fragments FAB de 4 IgG différentes contre des antigènes du VIH. Ces anticorps sont des immunoglobulines de Souris obtenues à la suite d'une immunisation avec des antigènes VIH. On ne peut pas les comparer avec les séquences des autres fichiers qui sont celles d'immunoglobulines humaines. En revanche, avec le logiciel *RasTop* par exemple, on peut visualiser la structure tridimensionnelle de ces immunoglobulines.

- ACY = IgG contre un épitope de l'antigène gp120 (protéine de l'enveloppe virale qui se lie au récepteur CD4 des cellules cibles du VIH ; cette protéine est codée par le gène « env » du virus) ;

- 1F58 = IgG contre un autre épitope de l'antigène gp120 ;

- 1E6J = IgG contre un épitope de l'antigène p24 (protéine formant la couche protéique interne du core du virus VIH ; cette protéine est codée par le gène « gag » du virus) ;

- 1NLD = IgG contre un épitope de l'antigène gp41 (protéine de l'enveloppe virale associée à la protéine gp120 et nécessaire à la fusion de cette enveloppe avec la membrane des cellules parasitées ; cette protéine est codée par le gène « env » du virus).

Documents fournis

Dans la banque de documents, le chemin **Immunologie/IgG/IgG anti VIH** permet de charger les fichiers :

- IgG.bmp qui présente la visualisation 3D obtenue avec le logiciel *RasTop* permettant de localiser les sites de fixation antigénique et de rappeler ce qu'est le fragment FAB ;

- igg.jpg qui révèle la structure tridimensionnelle d'une immunoglobuline ;

- vih.bmp qui affiche un schéma du VIH permettant de situer quelques protéines du virus qui constituent des antigènes contre lesquels sont produits des anticorps spécifiques (protéines P24, GP120, GP41 par exemple).

Fichiers des molécules en 3D

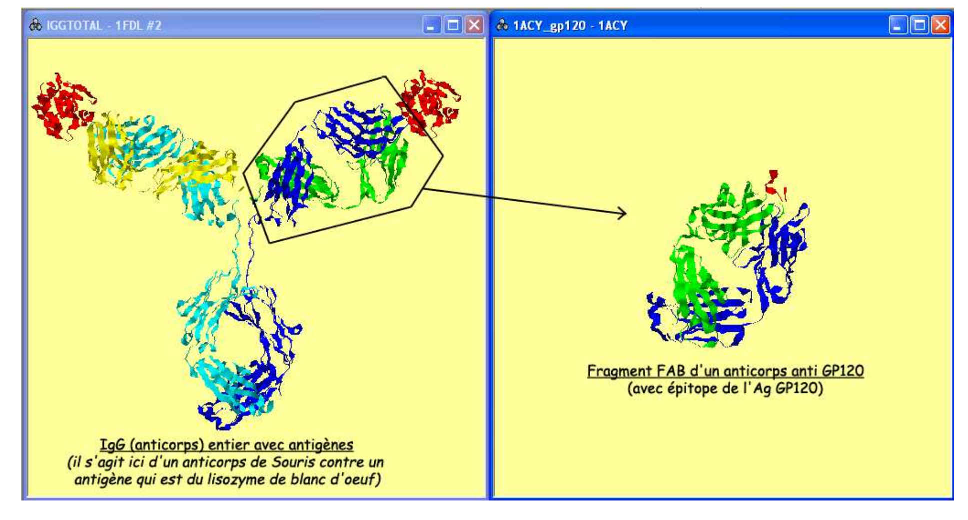
Dans le dossier 3D sous Anagene2, le répertoire **IgG** contient les fichiers .pdb suivants : 1ACY_gp120.pdb, 1E6J.pdb, 1F58.pdb, 1NLD.pdb, IGGTOTAL.PDB et LYS.PDB.

L'exploitation des données fournies permet d'expliquer la notion de spécificité des molécules d'IgG (anticorps). Les données fournies ici concernent des IgG dirigées contre des antigènes du VIH.

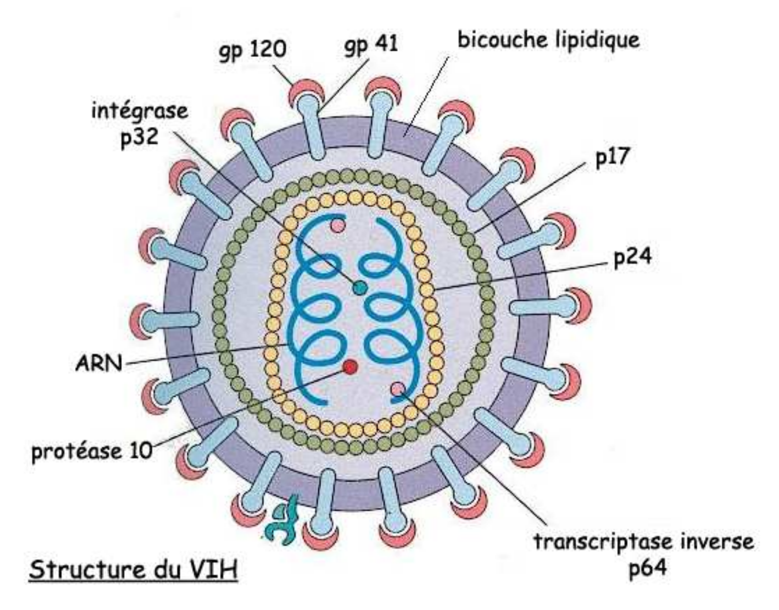
Prérequis

- Structure de la molécule d'IgG et localisation des sites de fixation antigénique.
- Mise en évidence de la spécificité des IgG.

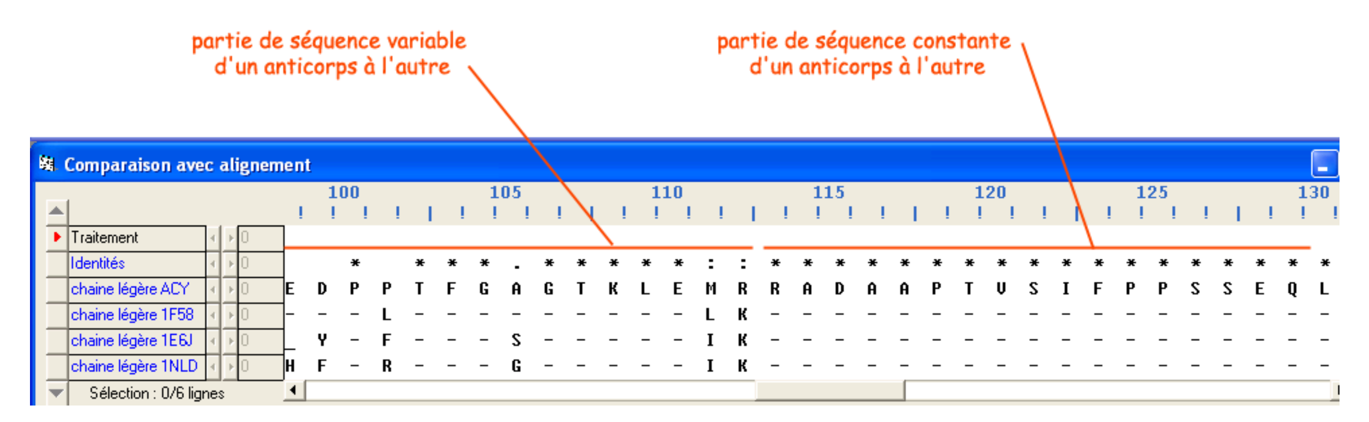
Le document IgG.bmp permet de rappeler la structure d'une molécule d'IgG, la localisation des sites de fixations antigéniques et de situer les fragments FAB (**ceux pour lesquels les séquences seront ensuite utilisées**).



Le document vih.bmp rappelle l'organisation du VIH et permet de situer les antigènes contre lesquels sont dirigés les IgG utilisés pour la démarche présentée.



La comparaison des chaînes lourdes et légères de quatre IgG différentes va permettre de mettre en évidence les différences qui existent entre les chaînes légères d'une part et les chaînes lourdes d'autre part.



Résultat obtenu par alignement avec discontinuité de quatre chaînes légères (fragments FAB seulement) de quatre IgG différentes

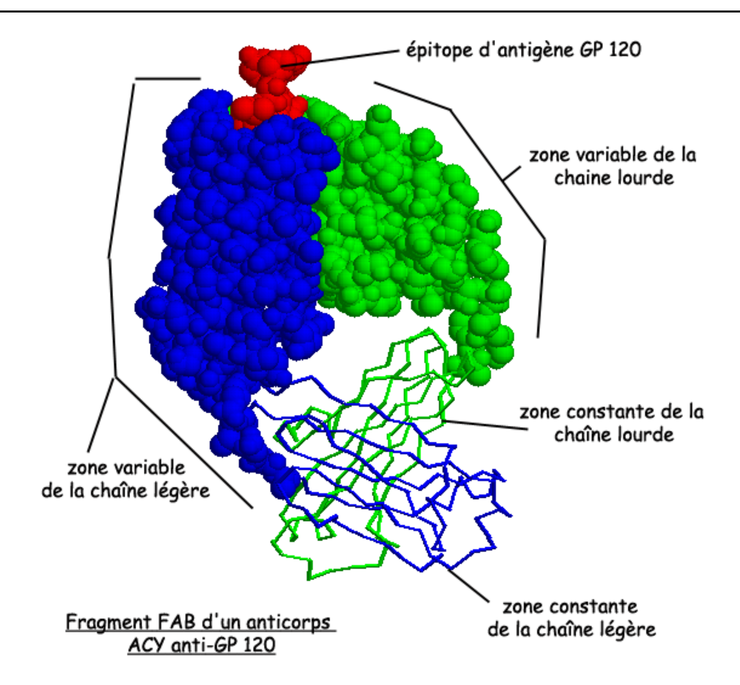
On peut ainsi constater de nombreuses variations d'un IgG à l'autre au début des séquences (acides aminés 1 à 112), mais aucune variation dans le reste de la séquence. On dégage alors la notion de partie constante et de partie variable des chaînes légères.

Un travail similaire effectué pour les chaînes lourdes permet d'arriver au constat de l'existence d'une partie variable d'un IgG à l'autre (acides aminés 1 à 117) et d'une partie constante (malgré quelques petites variations tout de même…).

On peut alors faire l'hypothèse que la spécificité des IgG est liée à l'existence de ces zones variables des chaînes lourdes et légères. S'il en est ainsi, ces zones variables (et plus précisément hypervariables) doivent être impliquées dans le site antigénique.

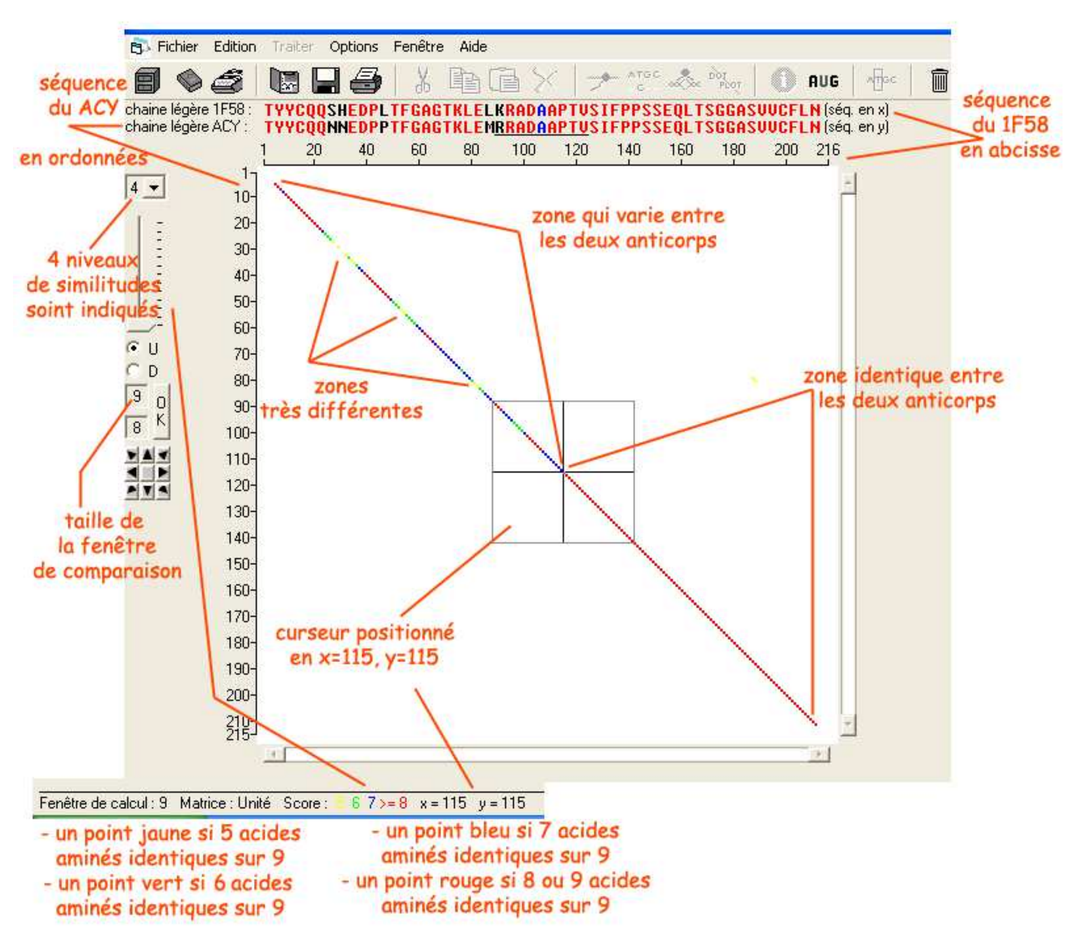
La localisation sur l'une de ces molécules d'IgG (ACY par exemple) des parties variables et constantes (avec le logiciel *RasTop*) permet de constater que les parties variables des IgG sont au niveau du site de fixation antigénique.

Le document igg.bmp montre le résultat que l'on peut obtenir avec le logiciel *RasTop* :



***Utilisation de la fonction DotPlot d'*Anagène**

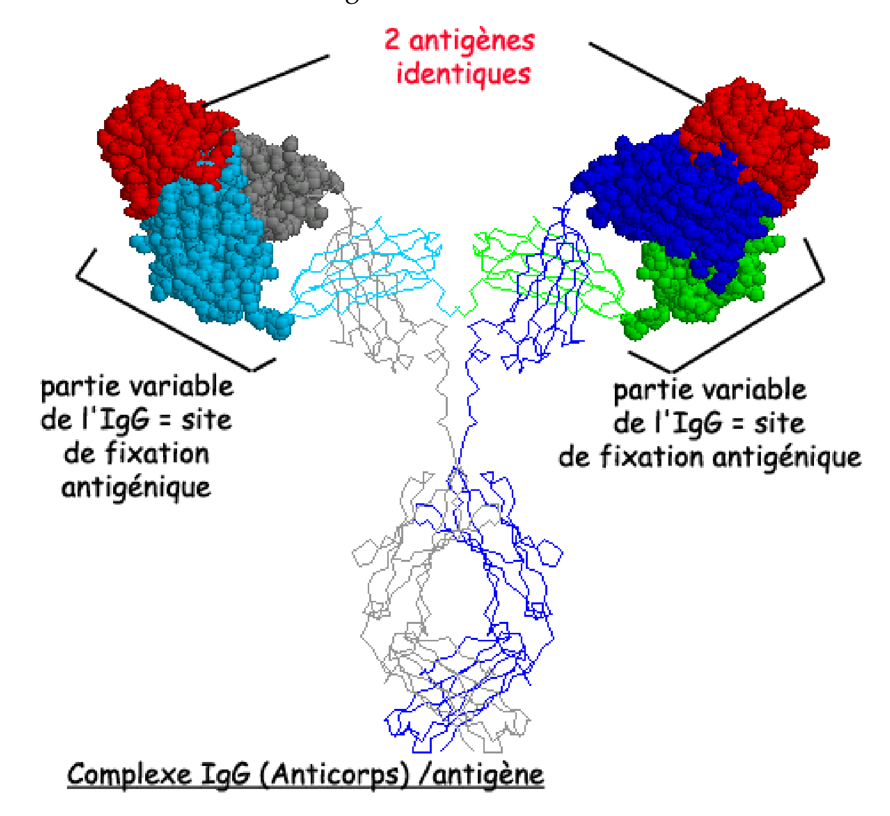
La comparaison des molécules d'IgG deux à deux peut aussi se faire avec la fonctionnalité DotPlot du logiciel. Cette fonctionnalité fournit un résultat plus « visuel » de la comparaison effectuée (qui équivaut à un alignement avec discontinuités) :



L'utilisation de cette fonctionnalité permet aussi de mieux repérer les sites hypervariables, au sein de la partie variable des IgG.

Bilan

On aboutit ainsi, en combinant l'exploitation des données fournies (pour *Anagène* et pour *RasTop*) à une représentation fonctionnelle de la molécule d'IgG :



Cette représentation est obtenue à partir d'un travail sur une IgG entière (IgG de Souris - fichier IgGtotal.pdb) : sélection et représentation en sphères des zones variables dans les quatre chaînes.
Il y a été ajouté les antigènes correspondant à cette IgG (lysozymes de blanc d'œuf – fichier lys.pdb).

Exploitation de données complémentaires

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Immunologie/IgG/Structure d'une IgG** permet d'atteindre :

- **Chaîne eu** qui charge le fichier igg-4chaines-eu.edi ;
- **Chaîne complète** qui charge le fichier 2igg-completes.edi ;
- **Chaînes légères régions variables** qui charge le fichier igg-variables-chaines-legeres.edi ;
- **Chaînes lourdes régions variables** qui charge le fichier igg-variables-chaine-lourde.edi.

Documents fournis

Dans la banque de documents, le chemin **Immunologie/IgG** permet d'atteindre **Structure d'une IgG** pour charger les fichiers IgG.bmp, igg.jpg et vih.bmp.

Les séquences précédemment utilisées sont limitées aux fragments fab.

On peut souhaiter débuter par une analyse de la molécule complète d'immunoglobuline. On dispose à cette fin d'autres données qui sont en rapport avec des anticorps humains autres que des immunoglobulines anti VIH.

L'analyse du document igg.jpg révèle que la molécule d'anticorps est formée de quatre chaînes, deux longues (lourdes), et deux plus courtes (légères).

Le fichier igg-4chaines-eu-edi fournit les séquences des deux chaînes légères et des deux chaînes lourdes d'une IgG humaine dirigée contre un déterminant antigénique d'un virus de la grippe. La comparaison des séquences des 4 chaînes de cette immunoglobuline révèle qu'elles sont identiques deux à deux, ce qui permet une schématisation plus précise de la molécules, sans toutefois fournir de réponse au problème de l'origine de la spécificité.

Le fichier 2igg-completes.edi fournit les séquences des chaînes lourdes et des chaînes légères de deux immunoglobulines humaines dirigées contre des déterminants antigéniques différents (l'un contre un déterminant antigénique d'un virus de l'hépatite, l'autre contre un déterminant antigénique d'un virus de la grippe). La comparaison des chaînes lourdes de ces deux IgG montre qu'elles présentent une région quasi identique (acides aminés 123 à 458), et une région comportant de nombreuses différences entre les deux molécules (acides aminés 1 à 122). De la même façon, la comparaison des chaînes légères de ces deux IgG montre l'existence d'une région quasi identique (acides aminés 100 à 215) et d'une région présentant de nombreuses différences dans la séquence d'acides aminés (acides aminés 1 à 99). On peut ainsi dégager la notion de région constante et de région variable d'un anticorps, et donc préciser encore le schéma de la molécule d'IgG. Cette conclusion peut être renforcée par la comparaison des chaînes légères et des chaînes lourdes de 4 molécules d'IgG anti-VIH (ici on compare donc 4 IgG dirigés contre 4 déterminants antigéniques différents d'un même virus).

Les fichiers igg-variables-chaines-legeres.edi et igg-variables-chaines-lourdes.edi contiennent les séquences des parties variables des chaînes légères et des chaînes lourdes de plusieurs molécules d'IgG humaines, dirigées contre des antigènes différents. L'utilisation du fichier igg-variables-chaines-legeres.edi permet une analyse plus fine des régions variables des chaînes légères et la reconnaissance de zones peu variables et d'autres très variables (hypervariables). Il en est de même de la comparaison des chaînes lourdes (fichier igg-variables-chaine-lourde.edi). Cela conduit à affiner encore plus la schématisation de la molécule d'anticorps en y localisant les zones hypervariables. Cette schématisation permet de réfléchir sur l'origine de la spécificité des anticorps et d'émettre l'hypothèse qu'elle est dépendante des zones hypervariables. Ces zones hypervariables peuvent aussi être mises en évidence avec les anticorps anti VIH de Souris.

Bases moléculaires de la spécificité spatiale des récepteurs T

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Immunologie/Récepteur des lymphocytes T** permet d'atteindre **Structure TCR** pour charger le fichier tcr-seq.edi affichant les séquences protéiques des chaînes alpha et bêta de 3 récepteurs T différents (1A07, 1BD2, 1J8H), reconnaissant des peptides viraux différents (pas forcément du VIH).

Documents fournis

Dans la banque de documents, le développement de l'arborescence **Immunologie** permet d'atteindre **Récepteur des lymphocytes T** pour charger le fichier recepteurT.bmp affichant la visualisation 3D d'un récepteur T complexé avec un peptide antigénique viral.

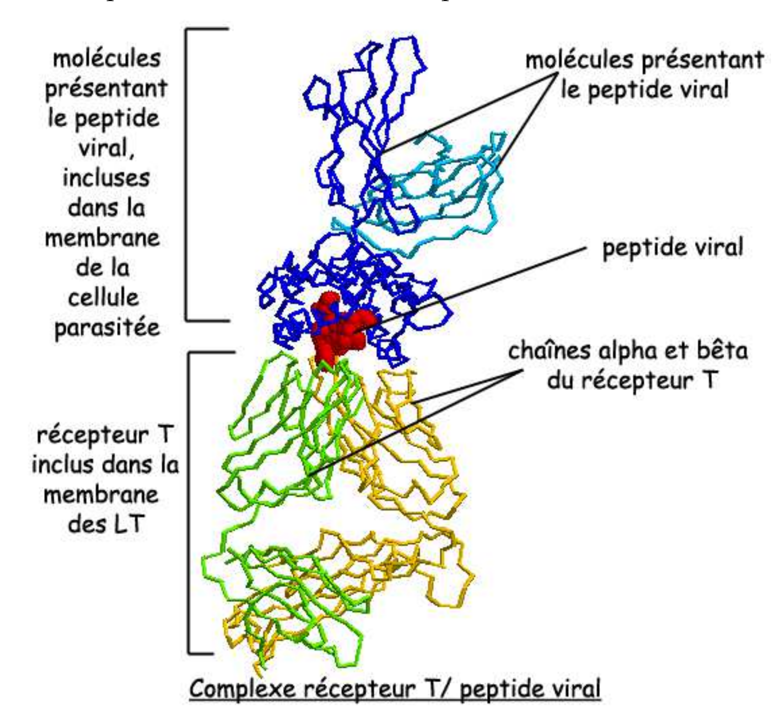
Fichiers des molécules en 3D

Dans le dossier 3D sous Anagene2, le répertoire **TCR** contient les fichiers .pdb suivants : 1a07.pdb, 1a07b.pdb, 1BD2.PDB, 1bd2b.pdb, 1J8H.PDB, 1j8hb.pdb, 1QSF.PDB et 1qsfb.pdb.

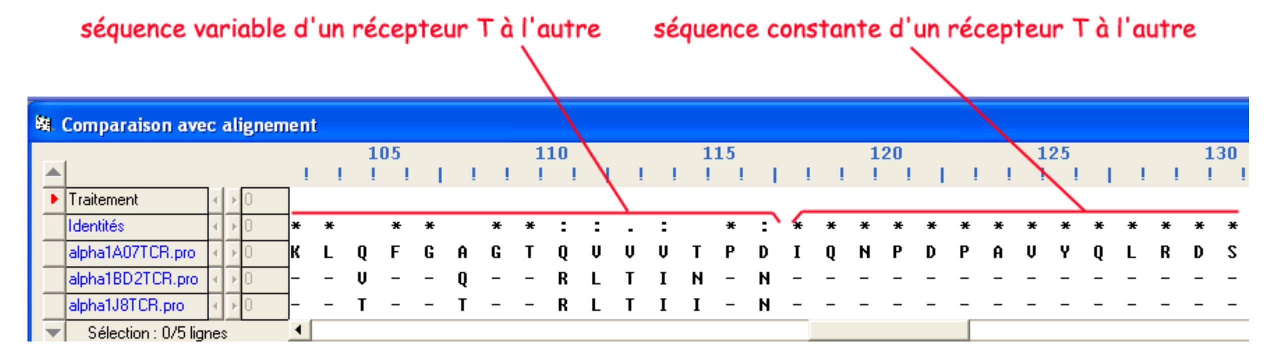
Pré requis

Spécificité des lymphocytes T grâce à leurs récepteurs T

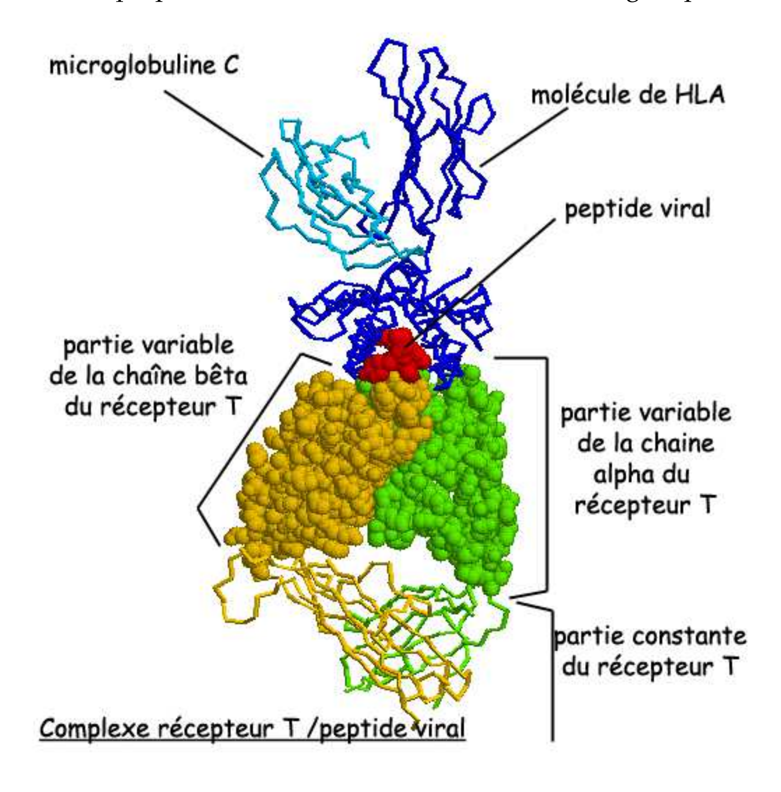
Le document recepteurT.bmp permet de présenter la structure d'un récepteur T et de constater que le contact avec le peptide viral ne se fait qu'à une extrémité de ce récepteur :



La comparaison des chaînes alpha entre elles et des chaînes bêta entre elles des trois récepteurs T différents permet de mettre en évidence une partie constante d'un récepteur T à l'autre et une partie variable :



Là aussi, l'hypothèse de l'implication de ces zones variables dans la spécificité des récepteurs peut être faite et testée. Avec *RasTop*, il est possible de localiser sur la structure d'un récepteur T les zones variables et voir si, au moins en partie, elles sont impliquées dans le site de reconnaissance antigénique.



Remarque

La comparaison des récepteurs T deux à deux peut aussi s'effectuer avec la fonctionnalité **Dot Plot** du logiciel *Anagène* (comme pour les IgG).

DES DÉBUTS DE LA GÉNÉTIQUE AUX ENJEUX ACTUELS DES BIOTECHNOLOGIES : *DIAGNOSTIC GÉNÉTIQUE*

 Albinisme

Informations scientifiques

Cf. page 76 (classe de première)

Pistes d'exploitation pédagogiques des données fournies

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Des débuts de la génétique aux enjeux actuels des biotechnologies/Diagnostic génétique/Albinisme** permet d'atteindre :

- **Polymorphisme tyrosinase et OCA2** pour accéder à :

Allèles tyrosinase qui charge le fichier all-tyrosinase.edi affichant les séquences strictement codantes de quelques allèles du gène de la tyrosinase ;

Allèles OCA2 qui charge le fichier all-OCA2.edi affichant les séquences strictement codantes de quelques allèles du gène OCA2 ;

- **Famille5 albinos l** pour accéder à :

Génotype tyrosinase qui charge le fichier AllelesTyrFamille5.edi affichant les séquences strictement codantes des allèles du gène de la tyrosinase de chacun des membres de la famille 5 et allèles de référence présents dans la famille (Tyrcod1, Tyrcod2, TyrAlbA3, TyrAlbA4). Les génotypes des individus sont les suivants : I 1, I 2, III 1 et III 2 sont (tyrcod1//tyrAlbA3) – I 3, I 4, II 1, II 3 et II 4 sont (tyrcod1//tyrcod1) – II 2 est (tyrAlbA3//tyrAlbA3) ;

Génotype OCA2 qui charge le fichier AllelesOCA2Famille5.edi affichant les séquences strictement codantes des allèles du gène OCA2 de chacun des membres de la famille 5, et allèles de référence présents dans la famille (OCA2Norm et OCA2m3). Les génotypes des individus sont les suivants : I 1, I 2, II 1, II 2 sont (OCA2 norm//OCA2 norm) – I 3, I 4, II 4, III 1, III 2 sont (OCA2 norm//OCA2 m3) – II 3 est (OCA2 m3//OCA2 m3).

Documents fournis

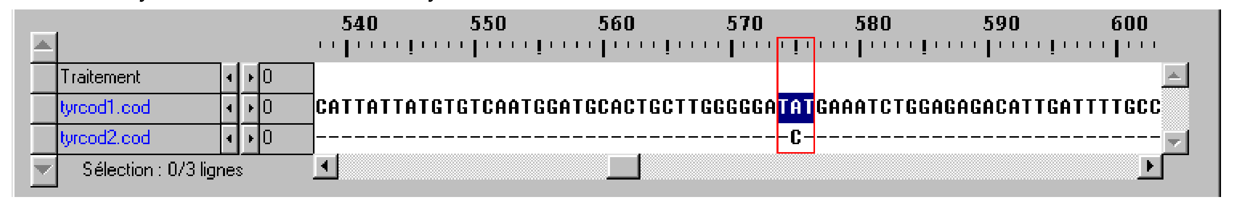
Dans la banque de documents, le chemin **Des débuts de la génétique aux enjeux actuels des biotechnologies/Diagnostic génétique** permet d'atteindre **Albinisme** pour charger les fichiers :

- arbre5.bmp qui présente l'arbre généalogique de la famille 5 ;
- electrophoreseSau3a.bmp qui montre la comparaison des électrophorèses pour l'allèle OCA2m3 et les autres allèles du gène OCA2, après action de l'enzyme de restriction Sau3a ;
- electrophoreseXbaI.bmp qui montre la comparaison des électrophorèses pour l'allèle tyralbA3 et des autres allèles du gène de la tyrosinase, après action de l'enzyme de restriction XbaI ;
- sau3aFamille5.bmp qui affiche l'électrophorèse de l'ADN des membres de la famille 5 après action de l'enzyme de restriction Sau3a ;
- Xba1Famille5.bmp qui affiche l'électrophorèse de l'ADN des membres de la famille 5 après action de l'enzyme de restriction XbaI.

Exercice préliminaire : principe de l'action des enzymes de restriction

Dans cet exercice préliminaire, il s'agit d'amener les élèves à s'approprier la technique d'identification d'un allèle grâce aux enzymes de restriction afin qu'ils puissent par la suite utiliser librement le logiciel pour déterminer le génotype d'un individu. Pour cela :

- charger les séquences Tyrcod1 et Tyrcod2
- comparer ces séquences et repérer l'existence d'une seule différence au niveau de la base 575 : A dans Tyrcod1 devient C dans Tyrcod2 ;



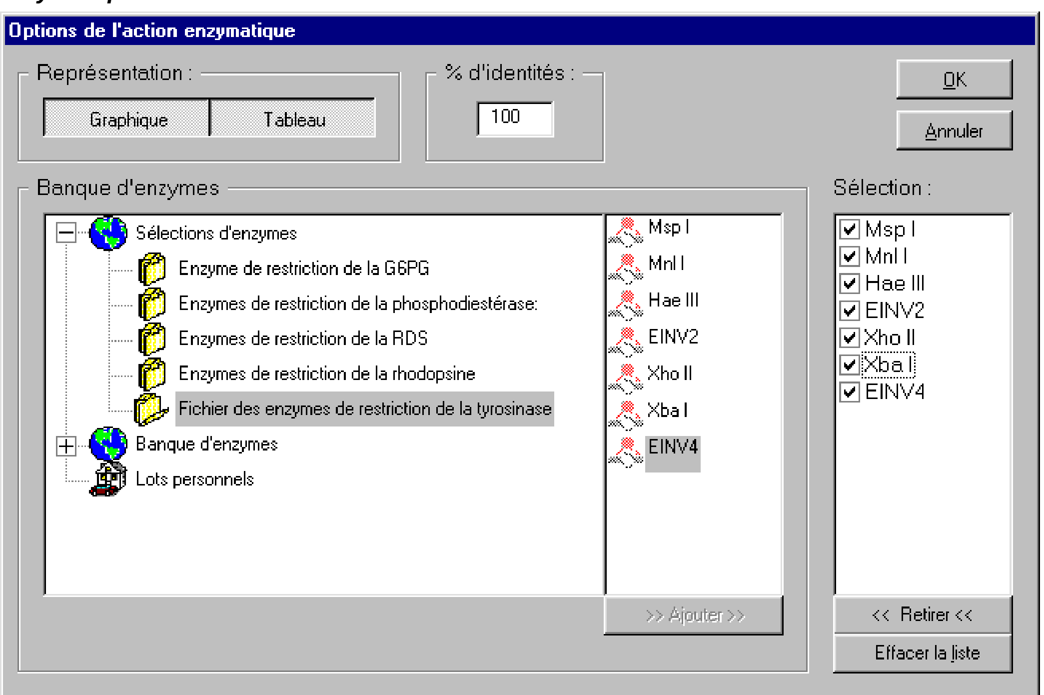
- introduire la notion d'enzyme de restriction et utiliser la banque des enzymes de restriction : repérer dans la liste des enzymes une enzyme susceptible de différencier les deux allèles Tyrcod1 et Tyrcod2. L'élève doit arriver à la conclusion que XhoII (GGATCT) coupe Tyrcod2 dans la région « 575 » mais pas Tyrcod1 ;

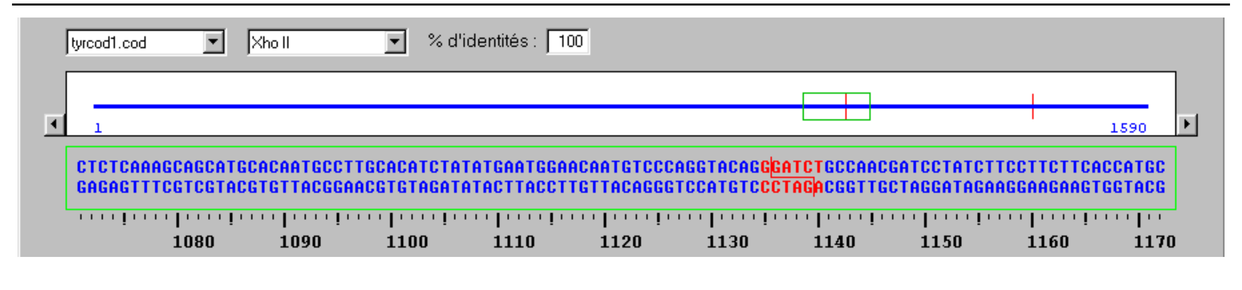
Enzymes du fichier **des enzymes de restriction de la tyrosinase**

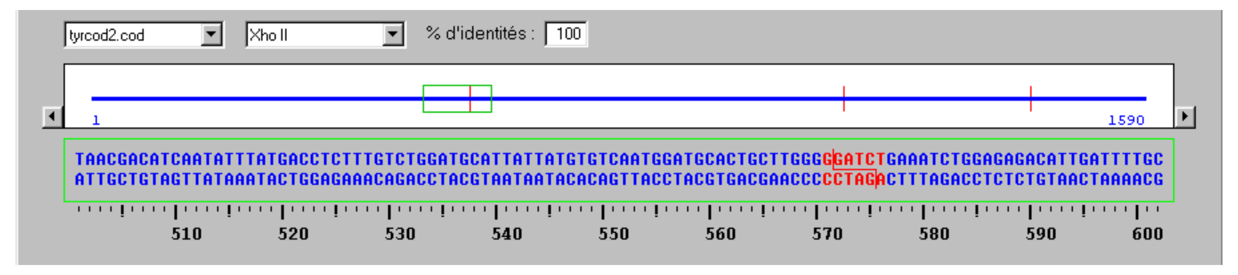
Nom	**Motif(s) reconnu(s)**
Msp I	CCGG
Mnl I	CCTC
Hae III	GGCC
EINV2	CAATC
Xho II	GGATCT
Xba I	TCTAGA
EINV4	AGTGTG

- vérifier cette conclusion en faisant agir cette enzyme et d'autres appartenant au fichier des enzymes de restriction de la tyrosinase sur Tyrcod1 et Tyrcod2.

Action enzymatique







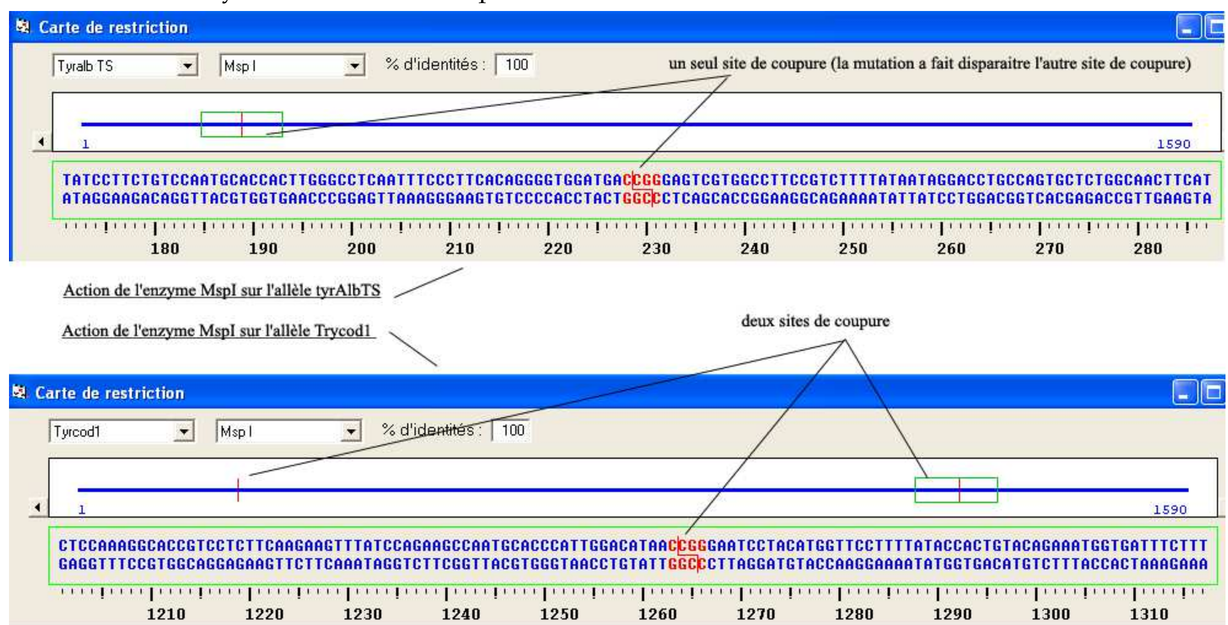
Visualisation de l'action enzymatique sous la forme d'un graphique. XhoII coupe deux fois Tyrcod1 (en 1135 et 1418) et 3 fois Tyrcod2 (en 571, 1135 et 1418). Le résultat obtenu signifie aussi qu'il y a plusieurs motifs GGATCT dans les séquences

Pour les autres enzymes, il n'existe aucune différence dans le nombre de coupures entre les deux séquences.

Enzymes (100%)	Msp I	Mnl I	Hae III	EINV2	Xho II	Xba I	EINV4
tyrcod1.cod	2	11	7	0	2	0	1
tyrcod2.cod	2	11	7	0	3	0	1
Sélect : 0/3 lignes							

Visualisation de l'action enzymatique sous la forme d'un tableau

- Un exercice du même genre peut être conduit pour distinguer les allèles Tyrcod1 et TyrAlbTS à l'aide de l'enzyme de restriction Msp1.



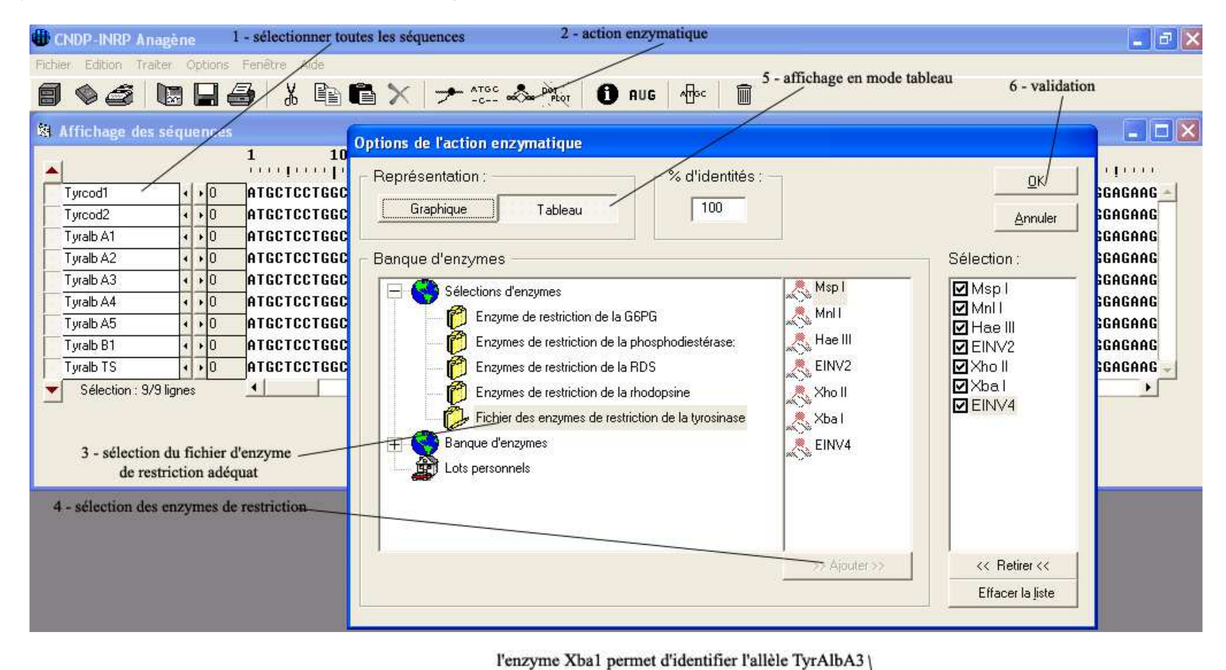
L'exemple ci-dessus présente l'action de l'enzyme de restriction MspI sur les allèles Tyrcod1 et TyrAlbTS

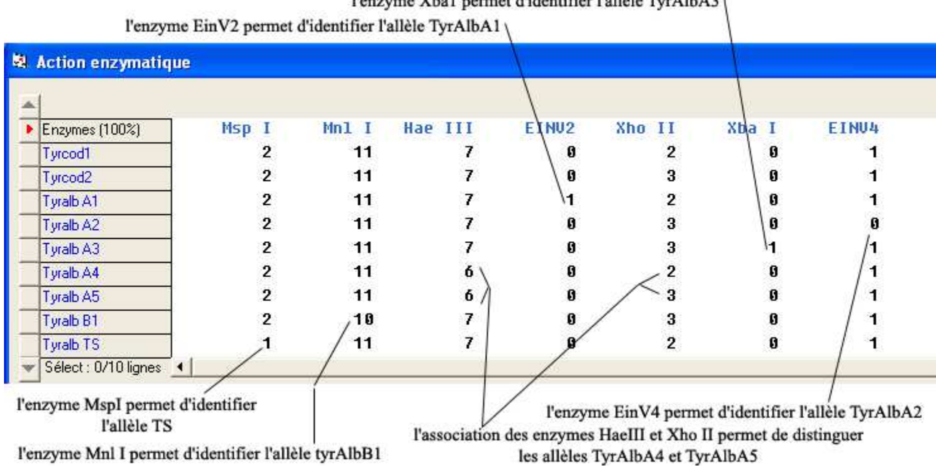
L'enzyme de restriction MspI reconnaît le site CCGG et coupe de façon dissymétrique au sein de ce site.

L'allèle Tyrcod1 comporte deux sites de coupure de l'enzyme (on obtient ainsi trois fragments de restriction de longueurs respectives : 0,228 kb, 0,035 kb et 327 kb) mais la mutation à l'origine de

l'allèle TyrAlbTS a fait disparaître le premier site de coupure (on n'obtient alors que deux fragments de restriction de longueurs respectives : 0,228 kb et 1,362 kb).

Pour identifier les enzymes de restriction qui vont permettre de reconnaître chacun des allèles du gène de la tyrosinase, il suffit de faire agir l'ensemble des enzymes de restriction sur les allèles de la tyrosinase et de demander un affichage des résultats sous forme de tableau.

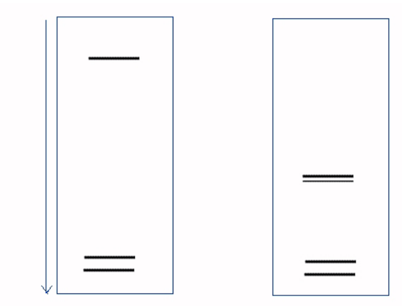




Enzymes (100%)	Msp I	Mnl I	Hae III	EINV2	Xho II	Xba I	EINV4
Tyrcod1	2	11	7	0	2	0	1
Tyrcod2	2	11	7	0	3	0	1
Tyralb A1	2	11	7	1	2	0	1
Tyralb A2	2	11	7	0	3	0	0
Tyralb A3	2	11	7	0	3	1	1
Tyralb A4	2	11	6	0	2	0	1
Tyralb A5	2	11	6	0	3	0	1
Tyralb B1	2	10	7	0	3	0	1
Tyralb TS	1	11	7	0	2	0	1

Cet exercice préliminaire permet ainsi de faire appréhender comment une enzyme de restriction spécifique permet de différencier deux allèles à partir du nombre de fragments obtenus à la suite de cette action.

On peut prolonger en indiquant que concrètement ces fragments peuvent être séparés par électrophorèse et visualisés à l'aide d'un révélateur de l'ADN. Le document peut être utilisé pour faire la liaison entre les données issues de l'utilisation de l'ADN et l'électrophorégramme.



Électrophorégramme résultant de l'action de l'enzyme de restriction XHO2 sur les allèles Tyrcod1 et Tyrcod2 qui codent tous les deux pour des enzymes fonctionnelles

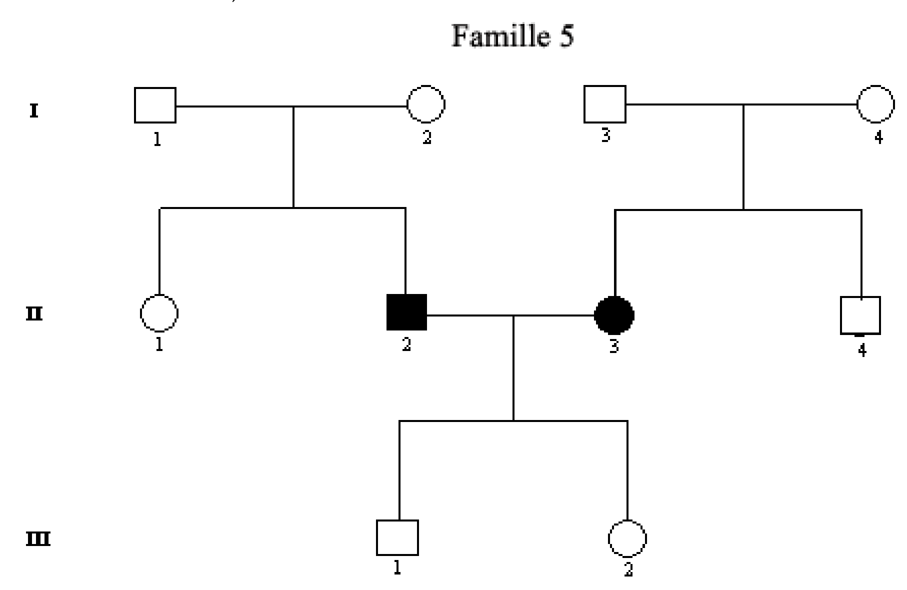
Analyse génétique d'une famille

La détermination des génotypes des membres de la famille 5 peut se faire de façon classique par comparaison des séquences des allèles du gène de la tyrosinase de chacun avec les allèles de référence fournis, mais les données permettent également à l'élève d'avoir une démarche se rapprochant de celle réellement suivie lors d'un dépistage génétique réel. C'est cette deuxième situation qui sera développée ci-dessous.

L'élève dispose :

- de l'arbre généalogique, qui lui permet de faire des hypothèses quant au déterminisme de la maladie et des prévisions pour les descendants ;
- des séquences des allèles du gène de la tyrosinase de chaque membre de la famille et des séquences de référence des allèles du gène de la tyrosinase présents dans la famille ;
- des résultats d'électrophorèse pour chaque individu de la famille (et d'électrophorèses de référence).

L'étude de l'arbre généalogique de la famille 5 permet de **faire l'hypothèse que deux gènes doivent intervenir** dans le déterminisme de l'albinisme (si l'on considère que les allèles mutés responsables du phénotype albinos sont récessifs) :



Phénotype de l'individu II2 :
- absence totale de pigmentation des cheveux, de la peau, des yeux ;
- troubles oculaires ;

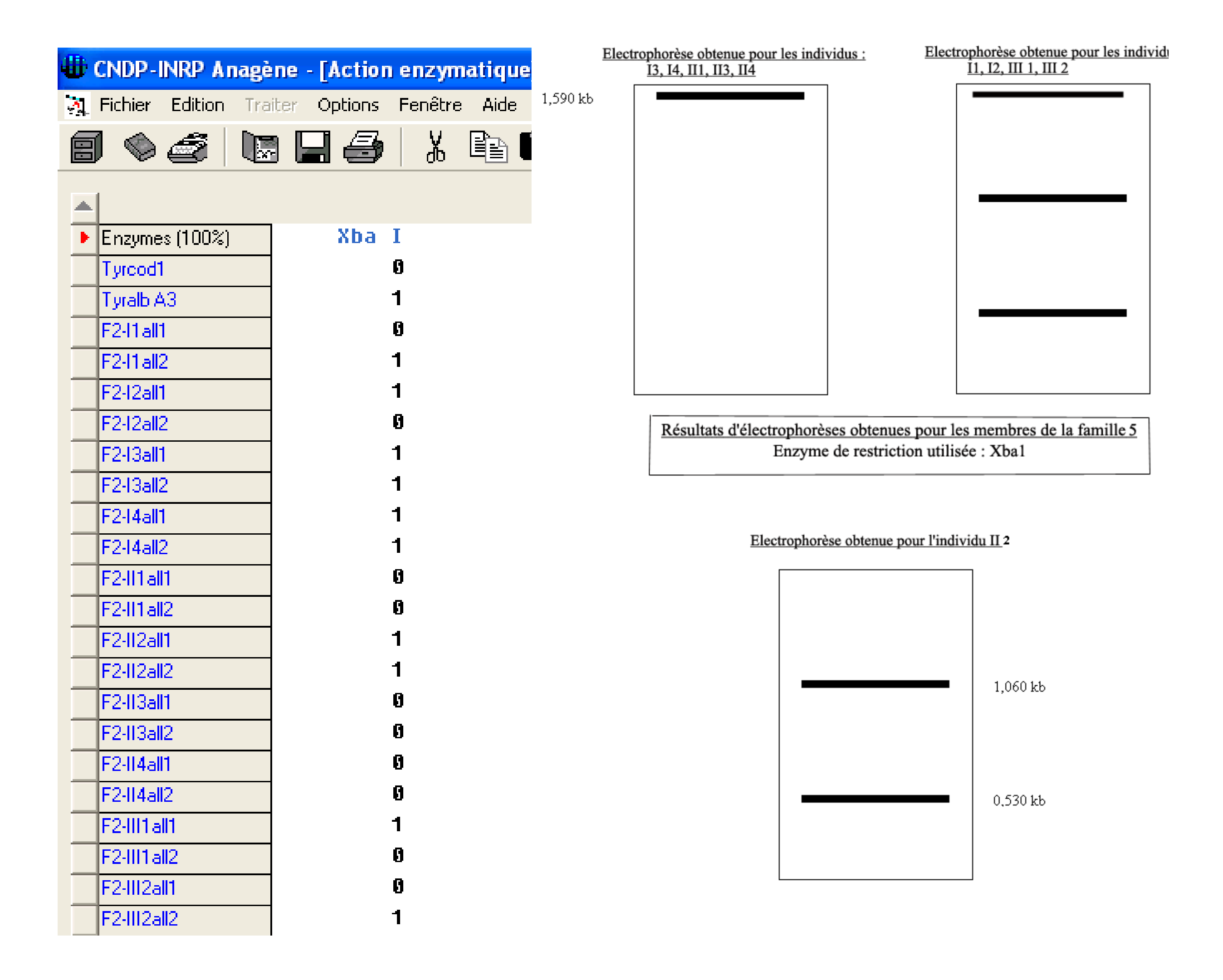
Phénotype de l'individu II3 :
- pigmentation des cheveux, de la peau, des yeux, faible à la naissance ; elle augmente durant la vie sans atteindre le niveau normal ;
- troubles oculaires ;

L'exploitation des résultats obtenus après action des enzymes de restriction (tableau et électrophorèses) permet de confirmer cette hypothèse et de préciser les allèles en cause.

L'exploitation des résultats fournis par l'utilisation des enzymes de restriction permet de déterminer les génotypes des membres de la famille 5 (on peut aussi, pour simplifier le travail, se limiter à la détermination du génotype des individus II2, II3, III1 et III2).

Détermination des génotypes pour le gène de la tyrosinase

Les allèles en jeu dans cette famille étant les allèles tyrcod1 et tyrAlbA3 (ce qui sera précisé aux élèves), on peut se limiter à l'utilisation d'une enzyme de restriction permettant de distinguer ces allèles : Xba1 par exemple. Les résultats du tableau sont à mettre en relation avec les électrophorèses correspondantes.



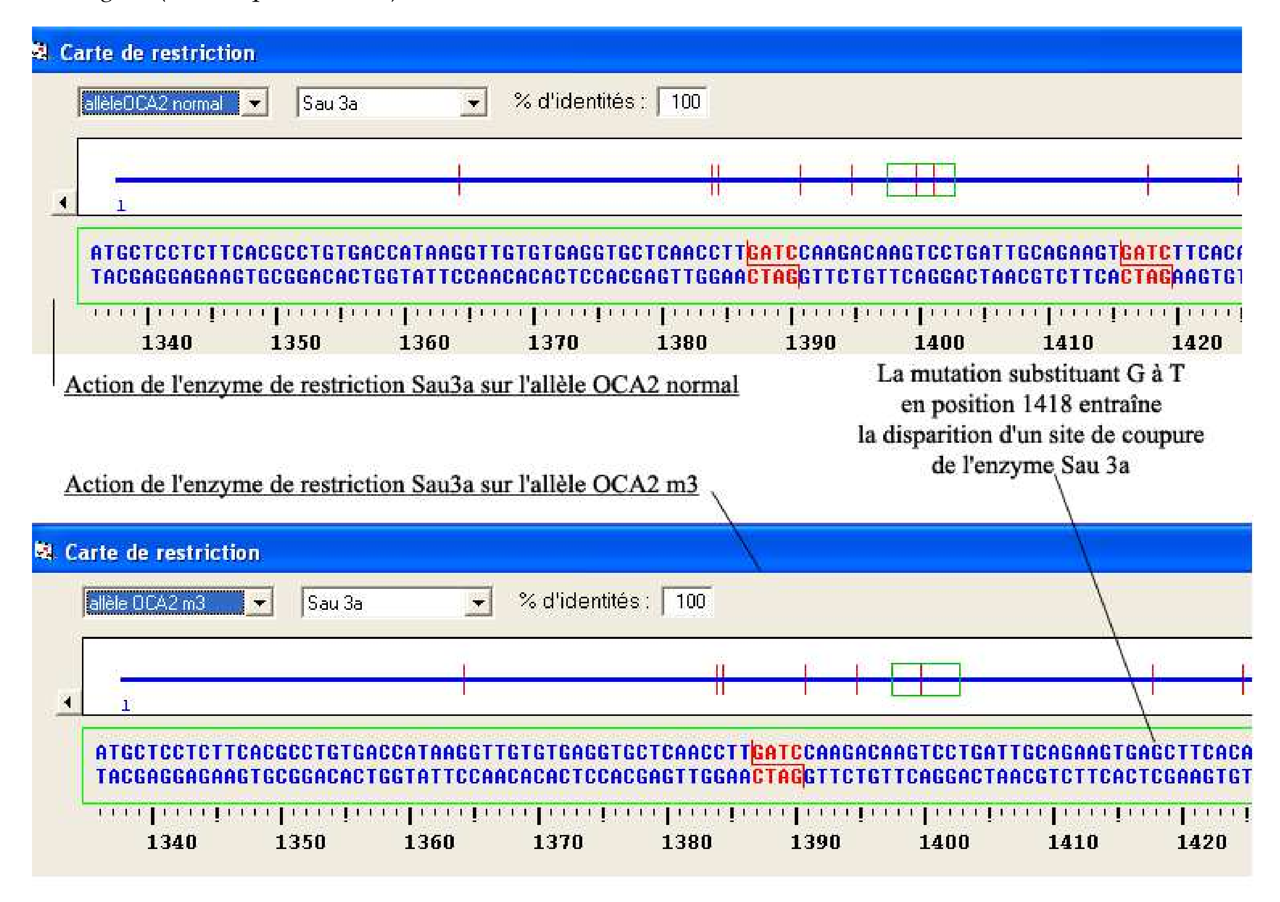
Enzymes (100%)	Xba I
Tyrcod1	0
Tyralb A3	1
F2-I1all1	0
F2-I1all2	1
F2-I2all1	1
F2-I2all2	0
F2-I3all1	1
F2-I3all2	1
F2-I4all1	1
F2-I4all2	1
F2-II1all1	0
F2-II1all2	0
F2-II2all1	1
F2-II2all2	1
F2-II3all1	0
F2-II3all2	0
F2-II4all1	0
F2-II4all2	0
F2-III1all1	1
F2-III1all2	0
F2-III2all1	0
F2-III2all2	1

Les électrophorèses seront interprétées en comparaison avec les électrophorèses de référence pour l'enzyme de restriction considérée

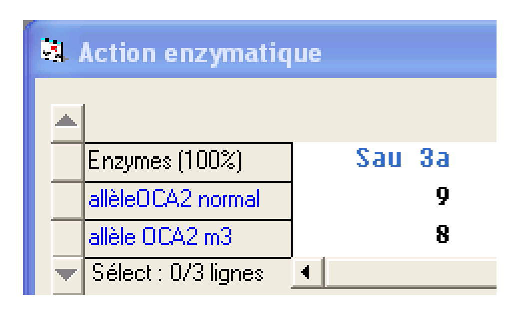
Les génotypes sont donc les suivants : pour l'individu II2 (tyrAlbA3//tyrAlbA3) – pour les individus I1, I2, III1 et III2 (tyrcod1//tyrAlbA3) – pour les individus I3, I4, II1, II3 et II4 (tyrcod1//tyrcod1).

Détermination des génotypes pour le gène OCA2

L'allèle en cause dans la famille 5 est l'allèle OCA2m3. Cet allèle diffère de l'allèle OCA2 par une substitution T –G en position 1418, ce qui supprime un site de reconnaissance de l'enzyme Sau3a (elle coupe en amont d'un site GATC). Cette enzyme de restriction Sau3a figure dans la banque d'enzymes d'*Anagène* (sites à quatre bases).

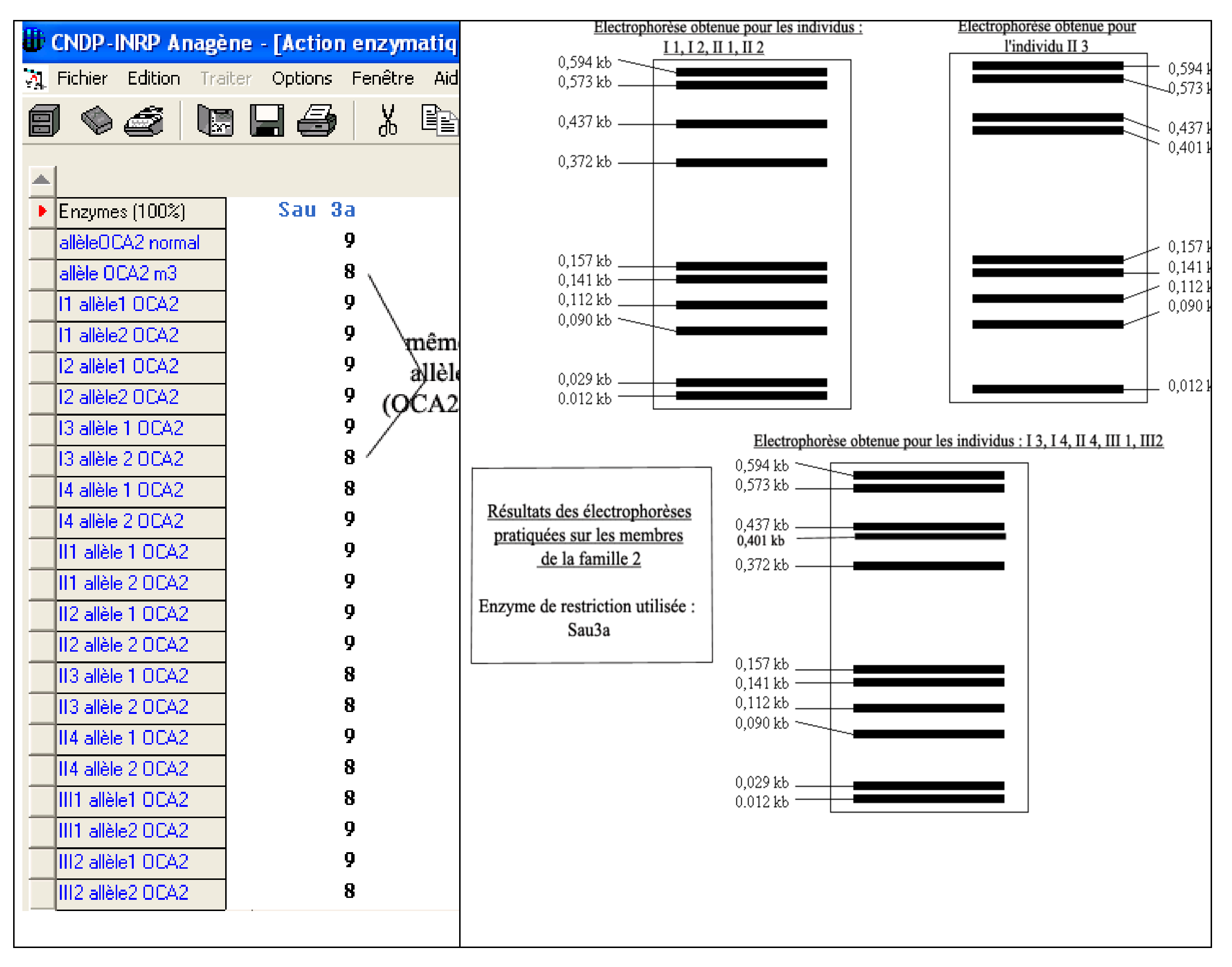


L'utilisation de cette enzyme de restriction permet donc de distinguer les allèles OCA2normal et OCA2m3 :



Enzymes (100%)	Sau 3a
allèleOCA2 normal	9
allèle OCA2 m3	8
Sélect : 0/3 lignes	

L'exploitation des résultats obtenus par action de cette enzyme Sau3a sur les allèles du gène OCA2 des membres de la famille 5 et les électrophorèses correspondantes permettent de déterminer leur génotype pour ce gène.



Les électrophorèses seront interprétées en comparaison avec les électrophorèses de référence pour l'enzyme de restriction considérée

Les génotypes sont donc les suivants : pour l'individu II3 (OCA2m3//OCA2m3) – pour les individus I1, I2, II1 et II2 (OCA2norm//OCA2norm) – pour les individus I3, I4, II4, III1 et III2 (OCA2m3//OCA2norm).

BIBLIOGRAPHIE

- « Génome humain et médecine », *Pour la Science*, janvier-mars 2005.
- ALBERTS B. *et al*, *Biologie moléculaire de la cellule*, Flammarion, coll. « Médecine Sciences », 2004.
- ALBIN R., TAGLE D., *Genetics and Molecular Biology of Huntington's Disease. Trends in Neurosciences*, vol. 18, n° 1, 1995, p. 11-14.
- BEAUDET A. L., SCRIVENER C. R., *The Metabolic and Molecular Basis of Inherited Disease*, McGraw-Hill, New York, 2000, 8e édition.
- CASTEX F., DIDES J.-J., *Enseigner l'immunologie en terminale S*, CRDP du Languedoc-Roussillon, 2003.
- HARRY M., *Génétique moléculaire et évolutive*, Maloine, 2001.
- LECOINTRE G., LE GUYADER H., *Classification phylogénétique du vivant*, Belin, 2005, 2e édition.
- LUCHETTA P. *et al*, *Évolution moléculaire*, Dunod, 2005.
- ORLANDO L., *L'Anti-Jurassic Park. Faire parler l'ADN fossile*, Belin, 2005.
- PICQ P., *Au commencement était l'Homme. De Toumaï à Cro-Magnon,* Odile Jacob, 2003.
- RAVEN P. H. *et al*, *Biologie végétale,* De Boeck Université, 2000.
- RAYMOND M. *et al,* « Insecticid resistance in the mosquito *Culex pipiens* », *Genetica,* 2001, 112-113, 287-296.
- RAYMOND M. *et al,* « La résistance du Moustique *Culex pipiens* aux insecticides », *Medecine Sciences* n° 12, vol. 19, décembre 2003.
- ROSSIGNOL J.-L. *et al, Génétique. Gènes et génomes*, Dunod, 2000.
- STRACHAN T., « Génétique moléculaire humaine », Flammarion, coll. « Médecine Sciences », 2003, 2e édition.
- THOMPSON J.D., HIGGINS D.G., GIBSON T.J., « CLUSTAL W. : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice », *Nucleic Acids Res. 22*, p. 4 673-4 680, 1994.

Documents disponibles au CNDP

- *Enquête sur l'apparition de la vie*, HAUBERTIN F., LE VOT B., CRDP de Bretagne, 1995. VHS. Référence : 350C2920.
- *Infogène, ressources pour l'étude de l'information génétique*, CNDP, 2002. Cédérom. Référence : 755 A0300.
- *La Drépanocytose, une maladie héréditaire*, CNDP, 1992, VHS 27 min. VHS. Référence : 002 P6254.
- *La Terre, le Temps, le Vivant : histoire de la vie et évolution*, CORGNAC F., DELETTRE A., FOUCHER R., REBOUT D., CNDP, 2005. DVD vidéo. Référence : 755B0670.
- V*isualiser les molécules en 3D avec le logiciel* Rasmol, DUCAMP P., TAILLARD J., CRDP d'Aquitaine, 2001. Guide d'utilisation du logiciel et recueil d'exercices. Cédérom. Référence : 3309M006.

INSTALLATION ET DÉSINSTALLATION

Le logiciel *Anagène* ne peut fonctionner qu'après installation sur le disque dur d'un micro-ordinateur de type PC Pentium équipé du système d'exploitation *Windows 98* ou d'une version postérieure.

INSTALLATION D'*ANAGÈNE*

Lors de l'installation, cette nouvelle version 2 d'*Anagène* ne supprime pas automatiquement l'ancienne version 1 qui peut coexister sur le même micro-ordinateur ; pour désinstaller l'ancienne version 1, il faut exécuter le programme **Desinst.exe** présent sur la disquette 1.

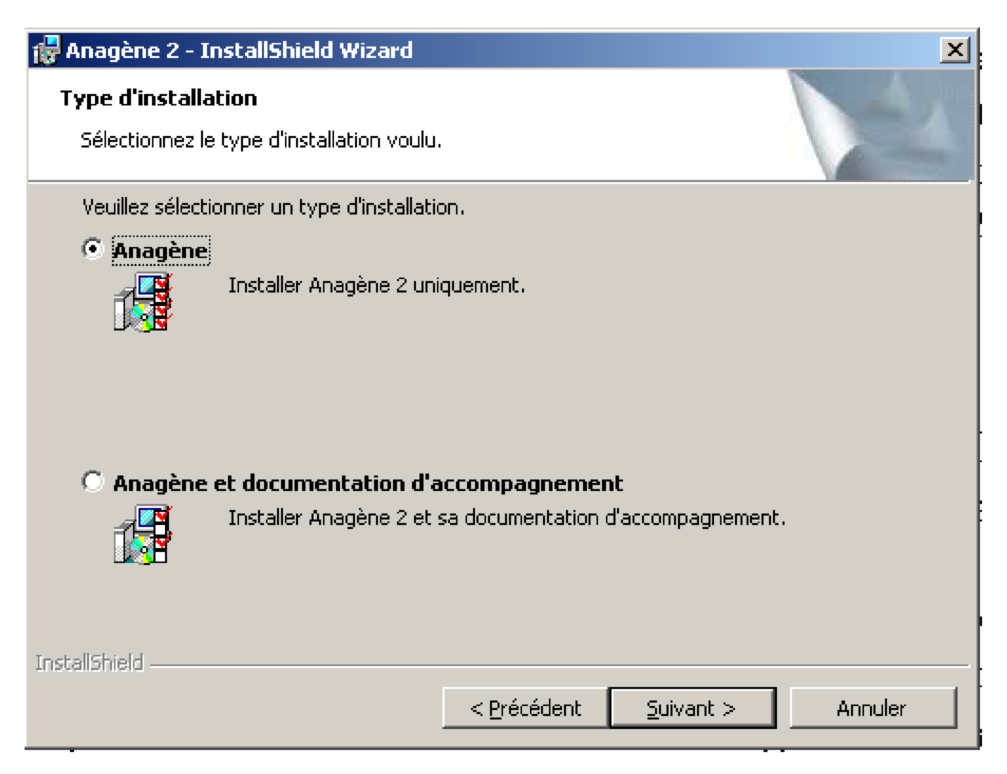
En revanche, la version 1.8 d'*Anagène*, si elle préexiste, est remplacée par cette version 2 sans qu'il soit absolument nécessaire de la désinstaller auparavant bien que cela soit néanmoins préférable.

L'installation sur un poste peut s'effectuer directement à partir du cédérom en local ou bien à partir d'une ressource en réseau (cédérom, disque dur d'un serveur de réseau ou d'une station, etc.).

Installation directe à partir du cédérom

Quelques instants après l'insertion du cédérom *Anagène* dans le lecteur, le programme d'installation s'exécute automatiquement (sous *Windows 98* et versions suivantes). Si le démarrage automatique n'a pas lieu, il faut pour lancer ce programme, après avoir cliqué sur **Démarrer** puis **Exécuter**…, entrer le chemin **D :\Anagene\Setup.exe** si D est la lettre désignant le lecteur de cédérom.

Après affichage de la fenêtre de bienvenue du programme d'installation d'*Anagène 2*, il suffit d'appliquer les consignes s'affichant successivement à l'écran. Deux types d'installation sont proposés :



La seconde option, si elle est sélectionnée, assure l'installation d'*Anagène* et copie de plus, sur le disque dur, la documentation d'accompagnement au format PDF. Elle nécessite, pour être lue à l'écran ou imprimée tout ou partie, la présence d'*Acrobat reader* fonctionnel dans sa version 4 ou postérieure. Une version de ce logiciel est disponible dans le dossier *Acrobat* du cédérom pour une éventuelle installation.

L'actualité des développements d'*Anagène* et des applications pédagogiques est l'objet des compléments en ligne accessibles sur les sites du CNDP et de l'INRP :

http://www.cndp.fr/svt/anagene/

http://www.inrp.fr/Acces/biotic/

Le fichier **Lisez-moi** situé à la racine du cédérom contient les informations d'assistance éditoriale sur ce produit.

Le cédérom *Anagène* n'est pas nécessaire lors d'une session de travail ; il sera rangé en lieu sûr pour une éventuelle nouvelle installation conformément aux clauses et conditions précisées dans la licence d'exploitation qui a été acquise.

Installation à partir d'une ressource en réseau

L'installation à partir du cédérom d'*Anagène* est également possible à distance en réseau si le lecteur d'une station, d'un serveur ou d'une tour de CD est au préalable partagé en lecture. Chaque station doit alors se connecter au lecteur de cédérom partagé et lancer l'exécution du programme **Setup.exe** situé dans le dossier **Anagene** du cédérom.

Une autre méthode d'installation consiste à copier au préalable l'intégralité dossier **Anagene** du cédérom sur une ressource en réseau, disque dur d'un serveur de réseau ou d'une station, etc. Ce dossier est à partager en lecture pour qu'il soit accessible en réseau à chacune des stations à équiper. Pour l'installation d'*Anagène*, chaque station doit se connecter à la ressource partagée et lancer l'exécution du programme **Setup.exe** situé à la racine du dossier. Après quelques instants, la fenêtre d'installation d'*Anagène* s'affiche au centre de l'écran de la station, indiquant les consignes à appliquer.

LANCEMENT D'*ANAGÈNE*

Anagène ne s'exécute pas à partir du cédérom, la présence du disque dans le lecteur n'est pas nécessaire. Pour améliorer le confort visuel, et si la carte graphique le permet, il est préférable de travailler avec une résolution d'au moins 800 x 600 pixels. On se référera au manuel d'utilisation de *Windows* ou à celui de la carte graphique pour effectuer les adaptations nécessaires.

Le lancement d'*Anagène* s'effectue en cliquant sur **Démarrer** puis **Programmes** puis sur l'une des deux icônes de lancement d'*Anagène* contenues dans le dossier **Anagène 2** :

- **Anagène** qui lance la version d'*Anagène* utilisant le code classique désignant par trois lettres chaque acide aminé d'une séquence peptidique ;
- **Anagène (code 1 lettre)** qui lance l'autre version d'*Anagène* utilisant le code international basé sur une lettre pour désigner les acides aminés d'une séquence peptidique.

Une troisième icône intitulée **Documentation** est présente si lors de l'installation, l'option correspondante avait été choisie.

DÉSINSTALLATION

Pour désinstaller *Anagène 2,* deux solutions sont possibles :

- utiliser le cédérom d'*Anagène* qui proposera, après lancement du programme d'installation, le choix entre trois options :

  - M**odifier les fonctions du programme** ;
  - **Réparer les erreurs dans le programme** ;
  - **Supprimer** ***Anagène*** **de votre ordinateur** ;

- afficher le **Panneau de configuration** de *Windows* et cliquer sur l'icône **Ajout/Suppression de programmes**, sélectionner **Anagene 2**, et cliquer sur **Supprimer**.

Attention, tout autre moyen de désinstallation (destruction de répertoires, effacement de fichiers...) est préjudiciable.